Medical Image Analysis 37 (2017) 22–36

Contents lists available at ScienceDirect

Medical Image Analysis

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

A framework for analysis of linear ultrasound videos to detect fetal presentation and heartbeat

M.A. Maraci[a,*], C.P. Bridge[a], R. Napolitano[b], A. Papageorghiou[b], J.A. Noble[a]

[a] Department of Engineering Science, Institute of Biomedical Engineering, University of Oxford, Oxford, UK
[b] Nuffield Department of Obstetrics and Gynaecology, John Radcliffe Hospital, University of Oxford, Oxford, UK

ARTICLE INFO

Article history:
Received 3 December 2015
Revised 22 December 2016
Accepted 5 January 2017
Available online 10 January 2017

Keywords:
Ultrasound video
Fetal presentation and heartbeat
Machine learning

ABSTRACT

Confirmation of pregnancy viability (presence of fetal cardiac activity) and diagnosis of fetal presentation (head or buttock in the maternal pelvis) are the first essential components of ultrasound assessment in obstetrics. The former is useful in assessing the presence of an on-going pregnancy and the latter is essential for labour management. We propose an automated framework for detection of fetal presentation and heartbeat from a predefined free-hand ultrasound sweep of the maternal abdomen. Our method exploits the presence of key anatomical sonographic image patterns in carefully designed scanning protocols to develop, for the first time, an automated framework allowing novice sonographers to detect fetal breech presentation and heartbeat from an ultrasound sweep. The framework consists of a classification regime for a frame by frame categorization of each 2D slice of the video. The classification scores are then regularized through a conditional random field model, taking into account the temporal relationship between the video frames. Subsequently, if consecutive frames of the fetal heart are detected, a kernelized linear dynamical model is used to identify whether a heartbeat can be detected in the sequence. In a dataset of 323 predefined free-hand videos, covering the mother's abdomen in a straight sweep, the fetal skull, abdomen, and heart were detected with a mean classification accuracy of 83.4%. Furthermore, for the detection of the heartbeat an overall classification accuracy of 93.1% was achieved.

1. Introduction

There have been significant advances in the analysis of ultrasound images in the last decade due in part to increased image quality but also the introduction of modern machine learning into the medical image analysis field (Noble, 2016). Machine learning is arguably very well-suited to recognize sonographic patterns in ultrasound images, which can form the basis of image-based decision-making. By contrast, traditional biomedical image analysis methods can find the dropouts, shadows, and sonographic signatures characteristic of ultrasound images difficult to accommodate, as they are the mapping of anatomy through the ultrasound image formation process. The most successful traditional methods in the literature are model-based methods that use strong geometric models as priors to cope with missing boundaries and artefacts.

Our particular interest is in obstetric ultrasound. The majority of the image analysis literature in this area has focused on automation of fetal biometry measurement for the anomaly scan (taken at 18–22 weeks gestational age). See Challenge US (Rueda et al., 2014) for a recent challenge that looked at a variety of methods and their performances. The anomaly scan is an essential ultrasound screening examination recommended worldwide for the detection of fetal abnormalities and early fetal growth restriction (Tiran, 2005). During a scan, a skilled sonographer acquires and records a number of two dimensional (2D) images of key fetal structures in diagnostic planes, following a standardized clinical protocol (typically a minimum of 6 but often more than 20 images) (Salomon et al., 2011). The goal is to diagnose structural abnormalities and to acquire biometry measurements that are verified against fetal growth charts. Research has looked into automating biometry measurement. For instance, Carneiro et al. (2008) used a discriminative constrained probabilistic boosting tree classifier for the detection and measurement of head, femur and abdominal structures. In their framework the probabilistic boosting tree classifier was trained on a database of key structures, where the nodes of the binary tree are strong classifiers trained using AdaBoost. Rahmatullah et al. (2011b); 2011a) used Adaboost for anatomical object detection in 2D fetal abdominal ultrasound images, where their framework was designed to identify whether the correct abdominal landmarks required for a standard plane

* Corresponding author.
E-mail address: mohammad.maraci@eng.ox.ac.uk (M.A. Maraci).

http://dx.doi.org/10.1016/j.media.2017.01.003
1361-8415/

were present. Sun (2012) applied a graph-based approach for automatic detection of the fetal skull, where initially the shortest circular path was detected. An ellipse was then fitted to the shape for finding the skull boundary. Ponomarev et al. (2012) applied a multilevel thresholding approach combined with edge detection and shape-based recognition for segmentation of the fetal skull. Imaduddin et al. (2015) used Haar-like feature with AdaBoost to detect fetal skull and femur. They further applied a Randomized Hough Transform for making biometry measurements. Anto et al. (2015) used a Random Forest to segment a head contour in fetal ultrasound scans that were acquired with a low-cost probe. Perhaps the most similar work to our own is the work of Lei et al. (2015), where densely sampled RootSIFT features were extracted and encoded using Fisher vectors for automatic recognition of fetal facial standard planes.

Three dimensional (3D) ultrasound was introduced in the 1990s as a technology designed to improve clinical workflow. It aimed to replace multiple 2D acquisitions by a single 3D acquisition, followed by standard plane finding in the volume. However, manual standard plane finding is quite time-consuming. This has led to a number of methods being proposed for automated plane finding (Chykeyuk et al., 2014; Yaqub et al., 2015) and some commercial systems now have automated plane finding as an option. However, the images from a 3D acquisition have a different appearance to those of a 2D acquisition and hence can contain different diagnostic value. It remains to be seen whether this type of solution will become accepted clinically. Quantification of 3D fetal ultrasound has, however, shown some promising results. For instance, Yaqub et al. (2011) successfully used Random Forests to perform fetal femur segmentation from 3D ultrasound volumes. This framework was later extended to automatically detect local brain structures in 3D fetal ultrasound images (Yaqub et al., 2012). Namburete et al. (2015) used Regression Forests to estimate the gestational age of a foetus from sonographic signatures in the brain. In the latter case, the accuracy of the method in the third trimester was shown to be higher than the current clinical standard.

It is important to note, though not often discussed, that in both standard 2D and 3D fetal sonography screening a sonographer follows a standardized clinical protocol, which defines criteria for the plane definition - see for instance the ISUOG guidelines for standard plane criteria (Salomon et al., 2011). Standardized 2D planes of acquisition undergo specific quality control to ensure they meet a set of predefined criteria. Moreover, sonographers need to be specifically trained to be able to meet these standards, as training programmes have previously shown to improve measurement variability (Sarris et al., 2011) and image quality (Wanyonyi et al., 2014). We refer to this standardized protocol as a **constrained scan**[1] since all images should have a similar appearance and contain certain anatomical structures, i.e. their appearance is deliberately constrained. These constraints can sometimes assist automated image analysis - for instance in abdominal circumference (AC) measurement, clear visualization of the stomach bubble, umbilical vein and often the spine is expected - but importantly reduce the degrees of variability with respect to the appearance of a general ultrasound scan of the foetus. Constrained scans are widely used in clinical practice, and simplify the image analysis challenge. However they have a key limitation. Acquisition of constrained scans requires a skilled sonographer. For wider adoption of clinical ultrasound in medicine and for uptake of ultrasound in the developing world, the need to acquire constrained scans has to be relaxed in favour of much simpler scanning protocols that a non-expert can readily learn.

Encouraging results from observational studies demonstrated that trained and standardized healthcare workers in developing countries can perform as well as qualified sonographers in terms of measurements reproducibility (Rijken et al., 2009). An automatic video acquisition analysis could potentially help in training, standardization and quality control in basic obstetric ultrasound for evaluating the fetal presentation and viability. The simplest scanning protocol to learn would be a linear ultrasound video sweep as illustrated in Fig. 1a. In our work, we propose the use of this type of scan and name it a **predefined free-hand** acquisition protocol. A novice sonographer can readily be trained to acquire data of this type. It is the analysis of data of this kind that we consider in this article. The question is then what useful diagnostic information can be automatically analysed from such videos?

To place our work in perspective, Fig. 2 schematically summarizes how some of the current state-of-the-art literature in fetal ultrasound image analysis maps between the skill needed for acquisition and type of image interpretation and analysis (none, detection & localization, quantification). As can be seen, most image analysis literature is in the lower third of this graph (data acquired by a skilled sonographer). We have included the assisted free-hand works of Kadour and Noble (2009); Kadour et al. (2010); Brown et al. (2013), which use controlled mechanical movement of the probe or subject for elastography on the middle row. These methods generate visualization of ultrasound information and require a small amount of user input to guide probe placement.

In recent years, several methods have been proposed for automatic detection and localization of anatomical fetal structures from ultrasound videos. Linear Dynamical Systems (LDS) were used to localize structures of interest in an ultrasound video obtained from a phantom by Kwitt et al. (2013). In our own work Maraci et al. (2014b), developed independently at around the same time, a method that performed well on clinical ultrasound video sequences was proposed. In that work, the original video is broken into smaller sequences of shorter length, where all sub-sequences have the same length. The dynamics of the sequences are then learned using a linear dynamical system. Identification and classification of the sequences of interest are then based on the similarities between the estimated LDS model parameters.

In an attempt to automatically find the image best representing the fetal abdominal standard plane in a video sequence, Kumar and Shriram (2015) used a method based on the spatial configuration of key anatomical landmarks. In previous works on which the current paper builds, we have investigated the bag of visual words approach with feature symmetry filters (Maraci et al., 2014a) as well as improved Fisher vector (IFV) encoding (Maraci et al., 2015) with a support vector machine (SVM) to identify frames of interest in an ultrasound video.

Finally, CNNs are gaining popularity in medical image analysis including analysis of ultrasound images although they are best suited to very large datasets and balanced data (which we do not have in our application). Chen et al. (2015) used a convolutional neural network (CNN) for standard plane localization of the skull and abdomen from an ultrasound video although the details of acquisition were not stipulated. Gao et al. (2016) have recently used a CNN for partitioning ultrasound video and (Baumgartner et al., 2016) for standard plane detection. We discuss CNNs further in the Discussion section.

To the best of our knowledge, the automation of the task of detecting the fetal presentation and heartbeat from a "predefined free-hand" ultrasound video has not been attempted before. We propose a three-step detection framework for characterizing an ultrasound video obtained from a predefined free-hand constrained scan protocol for pregnancies beyond 28 weeks of gestation. The first step in our method automatically identifies the frames corresponding to the fetal skull, abdomen and the heart. This is used

[1] In the clinical setting this is referred to as a *standardized scan.*