



This is a repository copy of *A supervised texton based approach for automatic segmentation and measurement of the fetal head and femur in 2D ultrasound images.*

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/127444/>

Version: Accepted Version

Article:

Zhang, L., Ye, X., Lambrou, T. et al. (3 more authors) (2016) A supervised texton based approach for automatic segmentation and measurement of the fetal head and femur in 2D ultrasound images. *Physics in Medicine and Biology*, 61 (3). pp. 1095-1115. ISSN 0031-9155

<https://doi.org/10.1088/0031-9155/61/3/1095>

Reuse

Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

A supervised texton based approach for automatic segmentation and measurement of the fetal head and femur in 2D ultrasound images

Lei Zhang¹, Xujiiong Ye^{1*}, Tryphon Lambrou¹, Wenting Duan¹, Nigel Allinson¹, and Nicholas J. Dudley²

¹Laboratory of Vision Engineering, School of Computer Science, University of Lincoln, Lincoln. LN6 7TS, U K.

²United Lincolnshire Hospitals NHS Trust, Medical Physics, Lincoln County Hospital, Greetwell Road, Lincoln LN2 5QY, UK.

Abstract

This paper presents a supervised texton based approach for the accurate segmentation and measurement of ultrasound fetal head (BPD, OFD, HC) and femur (FL). The method consists of several steps. First, a non-linear diffusion technique is utilized to reduce the speckle noise. Then, based on the assumption that cross sectional intensity profiles of skull and femur can be approximated by Gaussian-like curves, a multi-scale and multi-orientation filter bank is designed to extract texton features specific to ultrasound fetal anatomic structure. The extracted texton cues, together with multi-scale local brightness, are then built into a unified framework for boundary detection of ultrasound fetal head and femur. Finally, for fetal head, a direct least square ellipse fitting method is used to construct a closed head contour, whilst, for fetal femur a closed contour is produced by connecting the detected femur boundaries. The presented method is demonstrated to be promising for clinical applications. Overall the evaluation results of fetal head segmentation and measurement from our method are comparable with the inter-observer difference of experts, with the best average precision of 96.85%, the maximum symmetric contour distance (MSD) of 1.46 mm, average symmetric contour distance (ASD) of 0.53 mm; while for fetal femur, the overall performance of our method is better than the inter-observer difference of experts, with the average precision of 84.37%, MSD of 2.72 mm and ASD of 0.31mm.

Keywords—Ultrasound, image segmentation, fetal head, fetal femur, textons, automatic fetal biometric measurements

*Corresponding author: Xujiiong Ye. School of Computer Science, University of Lincoln, LN6 7TS, UK. E-mail: xye@lincoln.ac.uk

1. Introduction

Obstetric ultrasound (US) imaging is used in daily clinical practice due to its non-invasive nature, low cost and real-time acquisition (Rueda et al. 2014). Accurate US fetal biometrics play an important role in establishing the gestational age (GA) of the fetus, estimating its size and weight, and identifying fetal growth patterns and abnormalities, etc. To some extent, the high quality of obstetrics health care relies on high quality US images and corresponding accurate biometric measurements (Dudley and Chapman 2002). Common fetal US biometric parameters include: Crown-rump length (CRL), biparietal diameter (BPD), occipital-frontal diameter (OFD), head circumference (HC), femur length (FL) and abdominal circumference (AC) (Loughna et al. 2009; Pemberton et al. 2010).

In conventional clinical workflow for obstetric ultrasound examinations, sonographers are required to perform measurements manually using facilities such as built-in track balls and electronic calipers. This is manually intensive and time consuming. In addition, manual measurements can lead to a large variance of accuracy depending on the skill levels of individual sonographers. To this end, automatic approaches for fetal biometric measurements are needed to remove operator-dependence and to improve the quality, reproducibility and time efficiency of fetal biometric measurements (Espinoza et al. 2013). This will potentially increase examination throughput, leading to more efficient and cost-effective obstetric ultrasound services.

However, US images are one of the most difficult modalities to work with (Rueda et al. 2014). This is due to the image formation mechanisms intrinsic to ultrasound imaging, including low contrast, noise (artefacts, speckle), fuzzy or missing boundaries, and the inconsistency of US image quality (Dudley and Chapman 2002). As such the automatic US image segmentation and measurement has remained a challenging task.

1.1. Previous work on fetal segmentation

US fetal anatomical structure segmentation is a key step in accurate fetal biometric measurements. Existing approaches for US fetal image segmentation include Hough transformation (Lu et al. 2005), morphologic operators (Shrimali et al. 2009), parametric deformable models (Jardim and Figueiredo 2005), active contour models (Yu et al. 2008), machine learning (Carneiro et al. 2008b; Namburete and Noble 2013). According to different fetal biometric parameters, the objects of interest requiring measurement in fetal US image include fetal head, femur, and abdomen. For instance, Lu et al. (2005) proposed a fetal head segmentation method to measure BPD and HC based on an assumption that the contour of fetal head can be approximated as an elliptical shape with parameters progressively estimated by the iterative randomized Hough transform. Shrimali et al. (2009) proposed an algorithm to detect femur contour in fetal US image based on morphological operators. Ciurte et al. (2012b) proposed an algorithm to segment and measure the fetal abdomen in US images, in which a US image is represented as a graph of image patches, the segmentation is implemented by a user-assisted

initialization and a continuous min-cut partition of the graph and fast minimization scheme. Carneiro et al. (2008b) proposed a method for automatic detection of fetal anatomical structure in US images using a constrained probabilistic boosting tree classifier, several objects of interest (head, femur, abdomen, whole fetus) in fetal US images are segmented. The method was further developed and integrated into Siemens software called “auto OB” (Carneiro et al. 2008a), this is the only semi-automated commercial system for fetal biometry (Rueda et al. 2014).

1.2. Our Approach

To address the challenges in the automatic segmentation of fetal anatomic structure in US images, the approach we adopt is inspired by a principle of human visual perception and related research on modelling simple cell function in the receptive fields of the visual cortex (Julesz 1981; Petkov 1995; Petkov and Kruzinga 1997). With this in mind, we propose a novel supervised texton based approach for accurate segmentation of US fetal head and femur.

The contributions of this paper are three-fold:

- 1) A novel filter bank is designed to extract texton features specific to US fetal anatomic structure.
- 2) A unified boundary detection framework for US fetal head and femur is introduced considering multiscale local brightness, and texture cues. This is based on a state-of-the-art algorithm for contour detection and segmentation. However, additional modifications have been made to adapt the method to our US fetal anatomic structure segmentation.
- 3) A learning-based method for contour identification, together with geometric constraints of the object shape, is used for the final segmentation of US fetal head and femur.

The rest of this paper is organized as follows. Section 2 describes our method in detail; Section 3 presents materials and our experimental results, this is followed by discussion and conclusions which are given in the final section.

2. Methods

The proposed method is based on the concept that the geometrical profile of the skull can be approximated by an ellipse and the femur as a bar-like structure, whilst their appearances are brighter than other objects in fetal US images. Our method employs the texton technique that represents texture as a spatial arrangement of texture primitives. Information acquired in supervised training is then used in an edge detection stage which forms a critical part of the segmentation process. The segmentation is implemented as two sequential tasks which include edge detection and recognition followed by an object fitting process. Finally, the fetal biometric measurements can be derived from the segmented region.

A nonlinear diffusion technique is first used to reduce the influence of speckle in fetal US images.

In order to detect the boundaries of anatomical structures, we propose a scheme to extract features from structures by considering multiscale local brightness, and texture cues. A support vector machine (SVM) classifier is then used to identify the true edges (e.g. fetal head or femur) from the edge cues.

Based on edge segments extracted using the above method, the task of segmenting the head and femur is converted into object fitting. Due to the different geometric characteristics for fetal head and femur, typically the skull segments are arc-shaped objects in US images; whereas the boundary of femur is approximately a linear structure. In this paper, different object fitting processes are used for the segmentation of the head and femur in US image. The framework of our method is illustrated in Fig. 1. In the following subsections, each stage is described in detail.

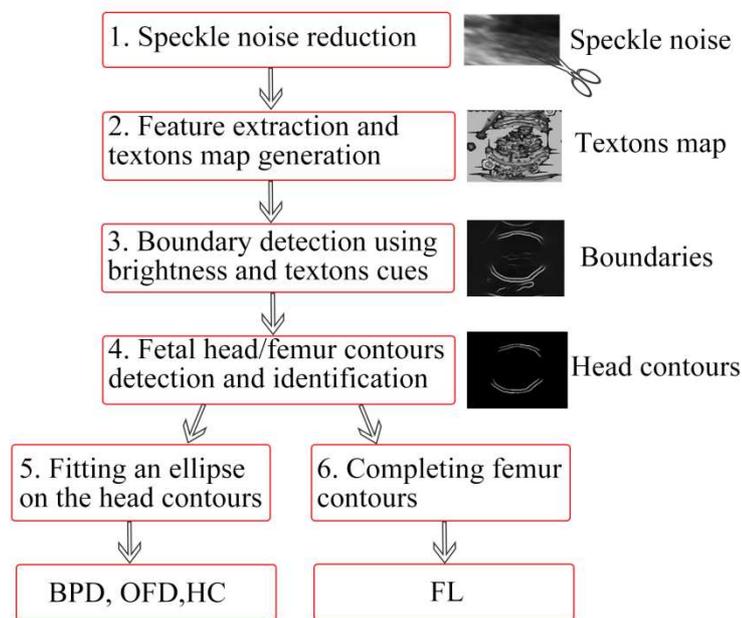


Figure 1: Block diagram of the proposed algorithm for head and femur measurements.

2.1. Speckle noise reduction.

Speckle noise is a signal dependent artefact that obscures clinically important details in US images due to the degradation of the contrast resolution between objects and the background. Speckle reduction is used as a pre-processing step for many US images processing tasks (Zhang et al. 2007). Among many speckle reduction techniques, nonlinear diffusion filtering (Weickert 1997b) is one of the most promising techniques that is commonly employed to remove speckle and enhance image quality. Perona and Malik (1990) proposed a nonlinear diffusion model which can selectively reduce noise and simultaneously preserve significantly important boundary information without blurring edges.

Let $u(\vec{x}, t): \Omega \rightarrow R$ be the grey scale image with a diffusion time t , for the image domain $\Omega \in R^2$, where $\vec{x}=(x_1, x_2)$ denotes image dimension, (here is two dimensions). The model is given by:

$$\begin{aligned}
 \partial_t u &= \text{div}((g(|\nabla u|^2) \nabla u)) && \text{on } \Omega \times (0, \infty) \\
 u(\vec{x}, 0) &= u_0(\vec{x}) && \text{on } \Omega \times (0, \infty) \\
 \partial_n u &= 0 && \text{on } \partial\Omega \times (0, \infty)
 \end{aligned} \tag{1}$$

where $u_0(\vec{x})$ denotes the original image as initial state of evolution, ∂_n is the derivative in the normal direction to the image boundary $\partial\Omega$. ∇u denotes the image gradient of Gaussian-smoothed version of u . To avoid smoothing at edges, the diffusivity $g(\cdot)$ is designed as a non-increasing function of the edge detector $|\nabla u|$, also known as a diffusivity function. Various forms of diffusivity function have been proposed. In our work, we employ a semi-implicit scheme called additive operator splitting (AOS) initially proposed by Weickert (1997a). The adopted diffusivity $g(\cdot)$ function is represented as the following form

$$g(\nabla u) = \begin{cases} 1 & (\nabla u \leq 0) \\ 1e^{\left(\frac{-3.315}{(\nabla u/\lambda)^4}\right)} & (\nabla u > 0) \end{cases} \tag{2}$$

where λ is a contrast parameter: the diffusivity $g(\nabla u) \rightarrow 0$ when structures with $|\nabla u| > \lambda$ are considered as boundaries, while $g(\nabla u) \rightarrow 1$ when structures with $|\nabla u| < \lambda$ are regarded as internal region of boundaries. Compared with other methods, the most pronounced advantage of Weickert's approach is that under typical accuracy requirements, the schemes are at least ten times more efficient than explicit schemes. More details can be found in (Weickert 1997a). Fig. 2 shows example images before and after speckle noise reduction processing. We can observe from Fig. 2 (b) that most of the speckle noise has been removed compared to the original image (Fig. 2 (a)). The image appears smoother, especially in the areas surrounding the skull while the most important head boundary information is preserved. This, in turn, benefits the feature extraction in the following texton generation stage and significantly reduces the false positive ratio of boundary detection, which will be discussed in the next section.

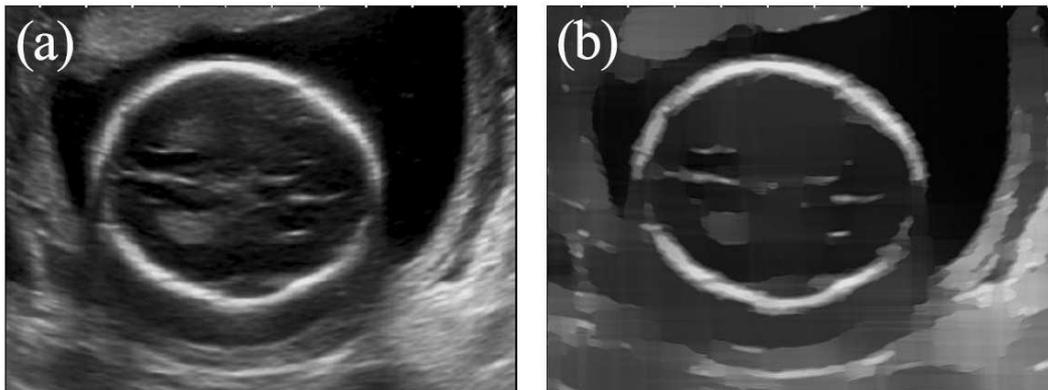


Figure 2 an example of speckle noise reduction for fetal head US image. (a) an original image and (b) the pre-processed image using nonlinear diffusion

2.2. Feature extraction and texton map generation.

Previous work in computer vision (Leung and Malik 2001; Martin et al. 2004; Varma and Zisserman 2004; Arbelaez et al. 2011) has demonstrated the advantages of using texture for providing significant information to distinguish the various patterns that present different visual features. For example, texture based segmentation methods in feature analysis, natural image segmentation and pattern recognition have been widely studied (Leung and Malik 2001; Martin et al. 2004; Arbelaez et al. 2011). Following this direction, we focus on investigating texton-based approaches. Texton is a powerful tool in texture analysis (Arbelaez et al. 2011). Texton was first introduced by Julesz (1981) in the 1980's but remained a vague concept (Zhu et al. 2005). It was then extended by Malik et al. (1999) and Leung and Malik (2001) to include an operational definition that a texture can be characterized by its responses to a filter bank (F_1, F_2, \dots, F_n) .

$$R=[F_1 * I(x,y), F_2 * I(x,y), \dots, F_n * I(x,y)] \quad (3)$$

Filter responses that are clustered into a set of prototype response vectors are defined as textons. This definition enables textons to be generated automatically from an image.

Filter based feature extraction approach is a common way to extract features used for texton generation. In our method, bar-like structure features, which represent bone structure (i.e. skull or femur) is extracted from US images. The cross-sectional intensity profile of bone structures can be approximated as Gaussian-like curves, while the intensities of those structures are on average higher than those of surrounding tissues. Fig. 3 (a) - (d) illustrates these profiles related to fetal head and femur, respectively.

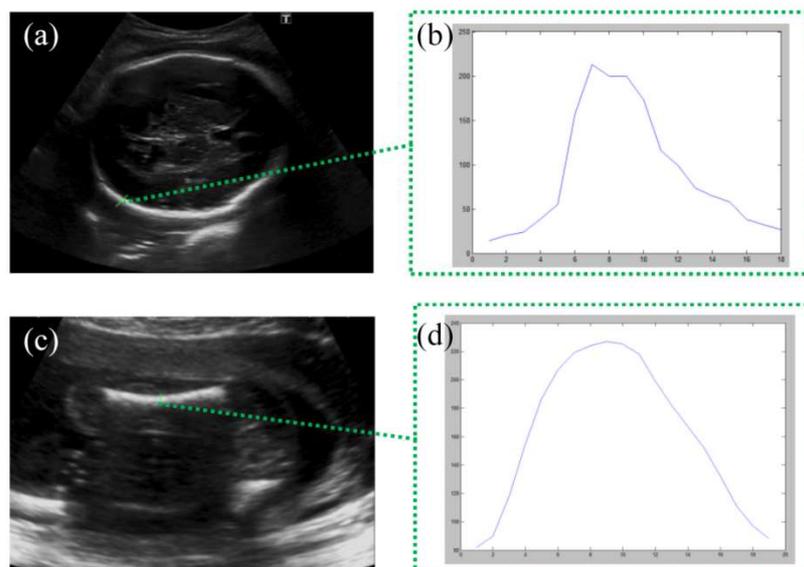


Figure 3: Cross-sectional intensity profiles of fetal head and fetal femur. (a) a fetal head US image sample with (b) cross-sectional profile of area indicated in (a) by green line. (c) a fetal femur US image sample with its cross-sectional profile (d).

In the work of Foi et al. (2014), the fetal skull is modelled using Difference of Gaussian (DoG). In our work, such structural properties are also taken into account when we design filter kernels to extract structure features

The design of a filter bank is an important part for texture generation. Different sets of filters have been investigated in the literature. Varma and Zisserman (2004) used a maximal response 8 (MR8) filter bank in their framework to classify natural texture patterns. In our method, we employ the second-order derivative Gaussian filter. Let us define two-dimensional Gaussian function as follows:

$$G(x,y)=\frac{1}{\sqrt{2\pi}\sigma_x}e^{-\frac{x^2}{2\sigma_x^2}}\times\frac{1}{\sqrt{2\pi}\sigma_y}e^{-\frac{y^2}{2\sigma_y^2}} \quad (4)$$

Given the skull or femur presented in US images may be rotated to any orientations, the filter is designed as an anisotropic filter kernel, that allows the kernel to be rotated, hence the rotated second order partial derivative of $G(x,y)$ with respect to the y -axis direction is given by

$$\frac{\partial^2 G(x',y')}{\partial y'^2}=\frac{1}{\sqrt{2\pi}\sigma_x}e^{-\frac{x'^2}{2\sigma_x^2}}\times\frac{1}{\sqrt{2\pi}\sigma_y^5}(y'^2-\sigma_y^2)e^{-\frac{y'^2}{2\sigma_y^2}}$$

$$x'=x\cos\theta-y\sin\theta \quad (5)$$

$$y'=x\sin\theta+y\cos\theta$$

A second order derivative Gaussian filter is applied at multiple scales (σ_x,σ_y) in order to cover different skull thickness or femur width in the US images and the anisotropic filter kernel at each scale is rotated in 12 orientations $(0^\circ, 15^\circ, 30^\circ, 45^\circ, 60^\circ, 75^\circ, 90^\circ, 105^\circ, 120^\circ, 135^\circ, 150^\circ, 165^\circ)$. These filter kernels are illustrated in the rows 1-3 of Fig. 4 (a). The surface of this kernel at one scale with corresponding cross sectional curve is illustrated in Fig. 4 (b). We also integrate the Matched Filter (Chaudhuri et al. 1989) into our filter bank to extract features. The Matched Filter was first proposed by Chaudhuri et al. to detect vessels in fundus images. The rotated matched filter kernel can be expressed by

$$P_i=[u\ v]=k(x,y)\times\begin{bmatrix}\cos\theta_i & -\sin\theta_i \\ \sin\theta_i & \cos\theta_i\end{bmatrix} \quad (6)$$

$$K_i(x,y)=-e^{-\frac{u^2}{2\sigma^2}}\ \forall P_i\in N \quad (7)$$

In which, the P_i denotes the points in a neighbourhood N defined in the area of $[u\ v]$. The i ($i=1, 2, 3 \dots, 12$) indicates the index of kernel which has a predefined angle. To eliminate the long double-sided tails of a Gaussian curve in N (in equation 7), the tails are truncated at $+3\sigma$ and -3σ , thus $|u|\leq 3\sigma$ (Chaudhuri et al. 1989). Meanwhile, $|v|\leq L/2$ is defined, in which L is used as the neighbourhood length of the kernel. The matched filter is then normalized to have zero mean as follows:

$$K'_i(x,y) = K_i(x,y) - \frac{1}{n} \sum_{P_i \in N} K_i(x,y) \quad (8)$$

where the number of points in N is denoted as n . In our experiment, we use a Matched Filter as a template to model the skull and femur at three scales. Same as the second order derivative Gaussian, the Matched Filter is also an anisotropic filter; so the kernel is rotated in 12 orientations for each scale. These filter kernels are illustrated in the rows 4-6 of Fig.4 (a). Fig. 4(b)&(c) illustrate these surfaces and cross-sectional profiles (shown inset in the top right). Additionally, we employ standard isotropic Gaussian filters (see Fig.4 (a) row 7) to extract general image features from the background.

For the anisotropic filters, at each scale, the maximal response, across all orientations, is considered. Therefore, there are in total 73 filter kernels in the filter bank, however only 7 responses (6 from anisotropic filter, 1 from isotropic filter) relate to the corresponding scales. The maximal responses of fetal head and femur across 12 orientations and the response to the Gaussian filter are illustrated in Fig. 5. These features are used to generate textons.

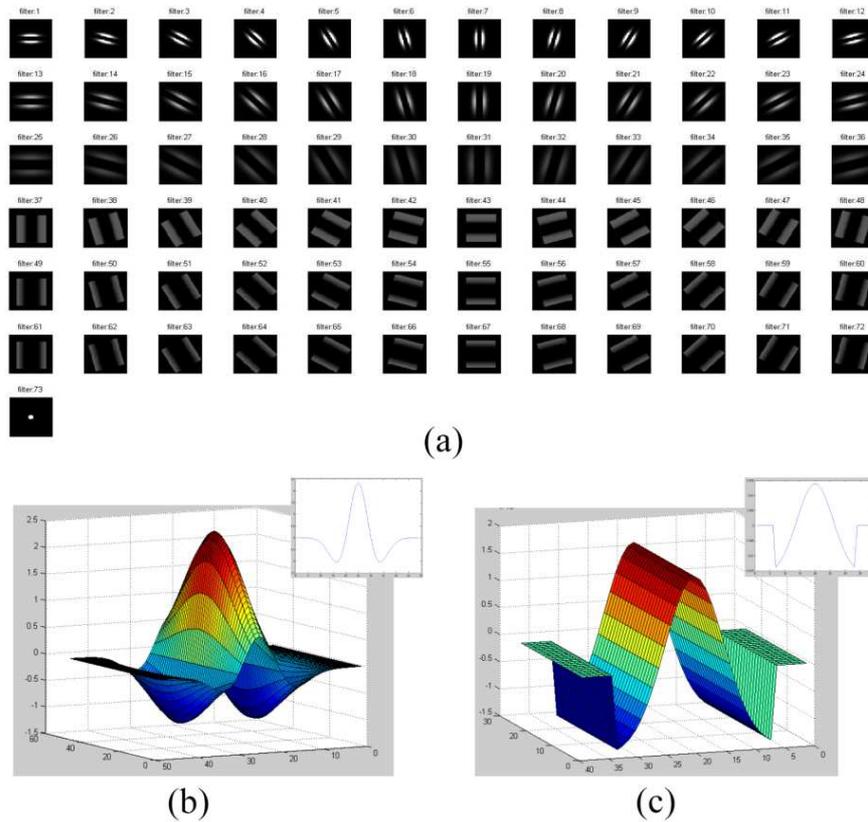


Figure 4 Filter bank used for fetal head and femur feature extraction. (a) Filter bank, first three rows are second order derivative Gaussian filter kernel at 3 scales ($\sigma=2,3,5$), rows 4-6 are matched filter kernel at 3 scales ($\sigma=2,3,5$), all of these anisotropic filters are rotated on 12 orientations which are illustrated in 12 columns. The standard Gaussian filter is illustrated at bottom row. (b) Second order derivative Gaussian filter surface with its cross sectional profile at scale $\sigma=3$ and (c) Matched filter surface with its cross sectional profile at scale $\sigma=3$.

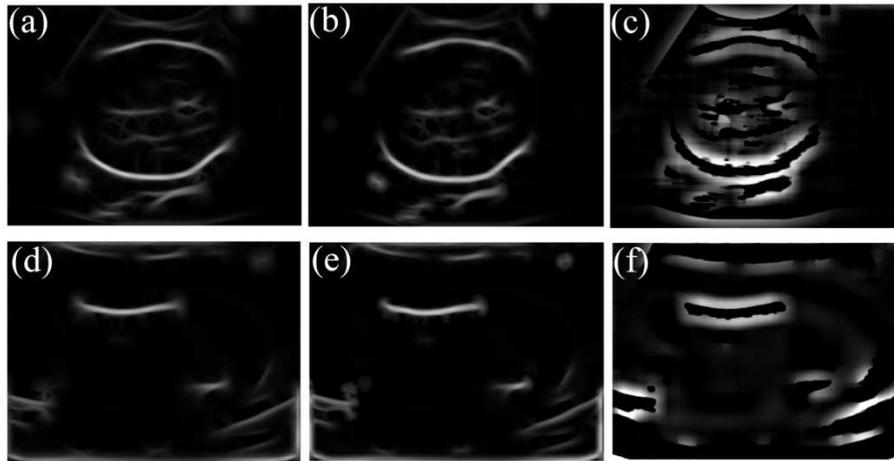


Figure 5. Examples of filter response: (a) Fetal head maximal response to the second derivative of Gaussian at scale $\sigma=5$ across 12 different orientations. (b) Fetal head maximal response to matched filter at scale $\sigma=5$ across 12 orientations. (c) Fetal head response to standard Gaussian filter at scale $\sigma=1$. (d) (e) (f) Femur responses related to second derivative of Gaussian, Matched filter and standard Gaussian, respectively.

Textons are generated from the filter bank responses, from which we extract local features of fetal head/femur in the US images. The textons were clustered by employing a *k-means* clustering algorithm on the filter responses. As representations of texture, the textons are aggregated based on the distances calculated from membership to clustering centers. The textons or texture primitives are linear combinations of the filters. Fig. 6 (c) demonstrates $k=32$ textons computed from a sample image. The number of clustering centroids ($k=32$) is chosen empirically according to the number of tissues that may present in the US images. For example, commonly, tissues in fetal head US image include, skull, brain, and other soft tissues, our experiments show $k=32$ is sufficient to generate corresponding primitives.

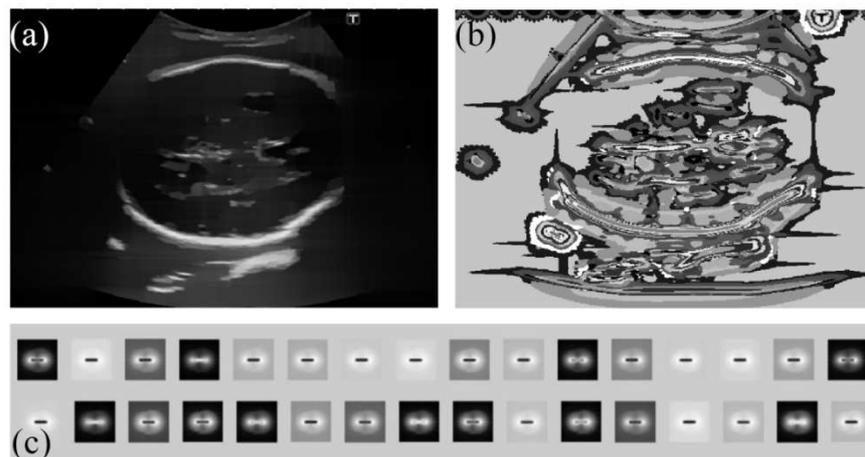


Figure 6. Textons generation: (a) is the pre-processed fetal head image (speckle noise reduction). The texton map shown in (b) is generated by assigning each pixel in the image (a) to the nearest texton. (c) textons dictionary which contains 32 textons calculated from training images.

Fig. 6 (a) is an example of the pre-processed image, Fig. 6 (b) is the associated texton map, which is generated by assigning each pixel in the image (Fig. 6 (a)) to the nearest texton, those pixels construct the texton memberships relating to each texton. Each texton is assigned a texton *id* using one of grey levels (from $k=1$ to 32). Therefore, the texton map is a grey scale image with values ranging from 1 to 32. The generated texton map (*Tmap*) is used in the next boundary detection stage, where the texture gradient is computed from the *Tmap*.

2.3. Initial boundary detection using brightness and texton cues.

Two steps are used to segment fetal head and femur in US images. The first step is head/femur boundaries detection. This is then followed by a closed contour construction. Recently, Martin et al. (2004) and Arbelaez et al. (2011) have proposed a contour detection method to accurately detect boundaries in natural images. Both local and global image information are combined to estimate the posterior probability of a boundary passing through the centre point. This allows some weak boundaries to be determined and excluded from significant boundaries while preserving contour quality. The method has been adapted in our work for accurate contour detection in an US image. Given the different conditions between the color (natural) image and gray level (US) image, we include brightness and texture features in our analysis instead of measuring the difference in local features in L,A,B (refer to CIELAB color space) and texture channels. A brief description of the method and how it is adapted to our US fetal anatomic boundary detection is given below:

The contour detection algorithm described in literature (Arbelaez et al. 2011) finds contours in the image by calculating an oriented gradient magnitude $G(x, y, r, \theta)$ at each location (x, y) of the image I . The gradients are generated in four channels (L , a colour, b colour and texture), separately, reflecting the local changes in brightness, colour and texture. A circular disc of radius r is initially drawn at a location (x, y) and splits it into two half discs along the diameter at orientation θ . The gradient magnitude is then calculated by employing χ^2 distance of histograms between two half discs (g and h) as:

$$\chi^2(g, h) = \frac{1}{2} \sum_i \frac{(g_i - h_i)^2}{g_i + h_i} \quad (9)$$

In our experiment, we calculate the oriented gradient signal in image brightness and texture channels based on Equation 9, where the brightness channel is the image intensity and texture channels is the generated textons map (*Tmap*). The linear combination of the local cues (brightness and texture) represents a single oriented signal, from which the general local cue is obtained by recording the maximum response across all pre-defined orientations θ :

$$mPb(x,y) = \max_{\theta} (\sum_i \beta_i G_i(x,y,\sigma_i,\theta)) \quad (10)$$

where i indicates feature channels (brightness or texture) and $G_i(x,y,\sigma_i,\theta)$ is histogram difference

function. The σ_i is the scale or radius parameter which determines disc size placed at location (x, y) for channel i . In our work, we considered the gradient at a single scale for each of brightness and texture channels. For the brightness channel, we use $\sigma = 9$ pixels, for texture channel we use $\sigma = 18$ pixels. The disc is split into two halves by a diameter ($2 * \sigma$) at orientation θ . The parameters β_i weight the contribution of each gradient signal. The weights β_i is trained by logistic regression fits using the training images.

Fig. 7 (a) shows an example of mPb . In the next step, global knowledge is incorporated by employing spectral clustering (Arbelaez et al. 2011). The local cues computed by applying oriented gradient operators at every location in the image are combined and yield a global eigenvalue problem. Specifically, the spectral clustering is calculated by constructing a sparse symmetric affinity matrix W , which encodes the maximal value of mPb along a line connecting two pixels, all pixels i and j are connected as a line segment \bar{ij} with a fixed radius r

$$W_{ij} = e^{(-\frac{\max_{p \in \bar{ij}}\{mPb(p)\}}{\rho})} \quad (11)$$

where ρ is a constant, here we use the same parameter values for r and ρ as in literature (Arbelaez et al. 2011), $r=5$ and $\rho=0.1$. The different eigenvectors $[v_0, v_1 \dots v_n]$ of W including their eigenvalues λ are then used to construct the spectral component of the boundary detector:

$$sPb(x, y, \theta) = \sum_{k=1}^n \frac{1}{\sqrt{\lambda_k}} \cdot \nabla_{\theta} v_k(x, y) \quad (12)$$

where $\frac{1}{\sqrt{\lambda_k}}$ is a weight parameter, and $\nabla_{\theta} v_k(x, y)$ is a gradient of the eigenvectors, which is generated by convolving each eigenvector v_k with anisotropic Gaussian derivative filters at predefined orientations θ .

The final global probability of boundary (denotes as gPb) is formed as a weighted sum of local signals mPb and spectral (global) signals sPb .

$$gPb(x, y, \theta) = \sum_i \beta_i G_i(x, y, \sigma_i, \theta) + \gamma \cdot sPb(x, y, \theta) \quad (13)$$

In this way both local boundaries cues and global image information can be integrated to extract more sophisticated boundaries while preserving contour quality. Fig. 7 shows examples of mPb , sPb and gPb maps extracted from a fetal head US image. As we can see from the Fig. 7, the appearances of weak boundaries presented in (a) and (b) which are related to the fetal brain are reduced in (c) while the head contours are preserved with a high probability level.

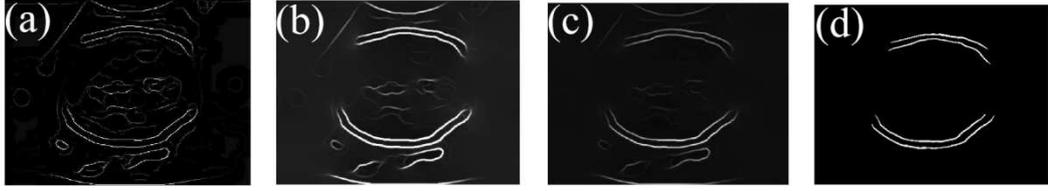


Figure 7 Examples of mPb , sPb and gPb extracted from a fetal head US image. (a) Local boundaries mPb , (b) Spectral boundaries sPb and (c) final globalized probability of boundary gPb . (d) Identified fetal head boundaries.

2.4. Fetal head and femur boundary identification.

Although the initial boundary detection discussed in the above section produces good quality contours in the US image, some non-head (or non-femur) boundaries are still present. For example, boundaries appear at the bottom of the image in Fig. 7 (c), which may influence ellipse fitting and the final segmentation. In the case of fetal femur, artifacts adjacent to the femur may present in US image, and those artifacts could appear similar to the femur in terms of brightness. To further identify the true head (or femur) boundary (or remove the false boundary), two steps are used: a) the global boundary probability map (gPb) calculated from Equation 13 is binarized using an optimal threshold which is computed based on grey level co-occurrence matrix (GLCM); and b) the objects (or contours) in the binary image are further classified into foreground (skull or femur) and background using support vector machine (SVM) classifier.

The GLCM (Haralick et al. 1973) is used to calculate the optimum threshold by computing the total second-order local entropy of the object and the background as

$$E_T^{(2)}(s) = E_A^{(2)}(s) + E_B^{(2)}(s) \quad (14)$$

where

$$E_A^{(2)}(s) = -\frac{1}{2} \sum_{i=0}^s \sum_{j=0}^s P_{ij}^A \log_2 P_{ij}^A$$

$$E_B^{(2)}(s) = -\frac{1}{2} \sum_{i=s+1}^{G-1} \sum_{j=s+1}^{G-1} P_{ij}^B \log_2 P_{ij}^B$$

where, p_{ij} is the probability of co-occurrence of gray levels i and j . The optimal threshold can be obtained by finding the grey level corresponding to the maximum of $E_T^{(2)}(s)$.

A boundary image is produced after thresholding the image gPb^* with an optimal threshold. In the next step, the SVM classifier is used to further remove non-skull or non-femur objects. A set of features is constructed for training the SVM classifier based on prior knowledge of the fetal skull and femur. These features include the pixel intensity, location, and shape feature. In our experiment, convex area, solidity of structure, major-, minor-axis lengths and eccentricity of an ellipse that best represents the structure, perimeter of the structure are taken into account for structural shape features

analysis. These features are calculated for each object in the binarized image and passed into the SVM classifier for training using the ground truths generated by experts. It is noted that, for both fetal head and femur, two different linear SVM classifiers are trained using 10-fold cross validation in the training stage. The trained classifiers are then used to correctly identify fetal head and femur structures. Fig. 7 (d) shows an example of the boundary identification using SVM, where some non-head boundaries have been removed.

2.5. Segmentation and measurements of fetal head and femur.

In general, the boundary detector does not produce closed boundaries. This is particular true for US image modality due to signal dropouts, the fetal skull is presented as an incomplete ellipse-like structure (see Fig. 6 (a)). In this section, for both fetal head and femur, two methods are used to construct closed boundaries from the identified boundary segments.

2.5.1. Segmentation and measurement of BPD, OFD and HC. The primary purpose of fetal head segmentation in US images is to obtain the standard biometric measurements: bi-parietal diameter (BPD), occipital-frontal diameter (OFD), head circumference (HC). Fig. 8 shows these measurements which are defined according to the real clinical application.

In our work, to construct a closed head contour, a direct least square ellipse fitting method (Fitzgibbon et al. 1999) is used to fit an ellipse to the skull boundaries obtained from the previous steps. The ellipse fitting process is implemented on an assumption that the fetal skull has an approximately elliptical shape (Foi et al. 2014). The core part of the ellipse fitting method is based on a least-squares technique, where the elliptical solution is resolved by minimizing the sum of squared algebraic distances from the data points (i.e. the skull boundaries) to the ellipse under an equality constraint. The optimal elliptical solution is computed directly using this method. An ellipse can be defined by an implicit second order polynomial:

$$F(x,y)=ax^2+bx+cy^2+dx+ey+f=0 \quad (15)$$

with an ellipse-specific equality constraint given by $4ac - b^2 = 1$, where a, b, c, d, e, f , are coefficients of the ellipse and (x,y) are coordinates of points lying on it. Please refer to (Fitzgibbon et al. 1999) for further details and theoretical demonstration.

In order to calculate BPD and OFD in an ‘outer to outer’ manner, the distances are measured between the outer borders of the identified skull. To ensure the fitted ellipse is based on the outer boundary of the identified skull, a maximal ellipse selection scheme is used. For example, there are two pairwise (four) skull segments (boundary) in Fig. 7, of which any two segments are selected to fit an ellipse. As a result, six different ellipses can be obtained based on different segment combinations, among which the maximal ellipse is considered. The BPD and OFD are then obtained by calculating minor axis length and major axis length from the fitted ellipse, respectively. The *HC* is calculated using the formula $HC=\pi(BPD+PF D)/2$.

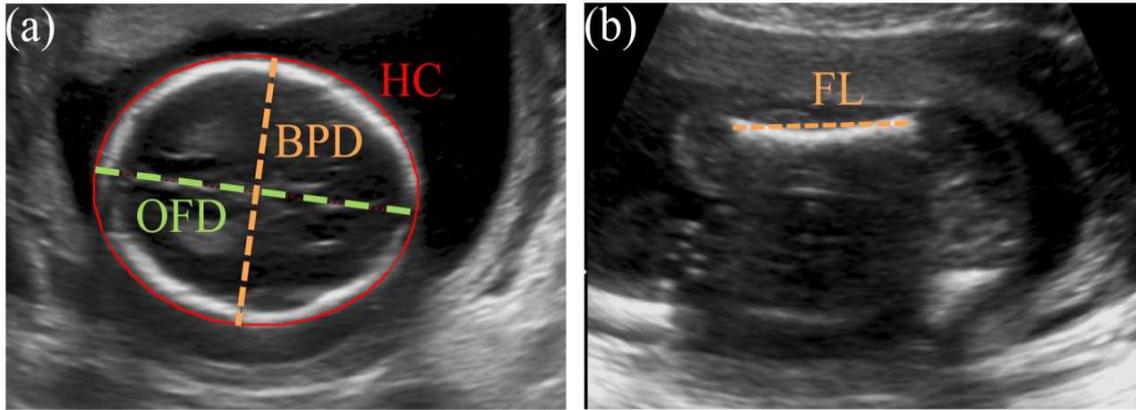


Figure 8. The biometric measurements of the fetal head (a) and femur (b) in US images. The BPD measurement is taken on the outer border of the parietal bones ('outer to outer') at the widest part of skull. The OFD is measured between the outer border of the occipital and frontal edges of the skull at the point of the midline ('outer to outer') across the longest part of skull. The HC is calculated from the formula $HC = \pi(BPD + OFD)/2$. FL in (b) is measured on the outer borders of the edges of the femoral bone.

2.5.2. Segmentation and measurements of FL. The same procedure is also applied to femur segmentation. However, instead of using an ellipse fitting method to produce a closed contour, we simply connect two femur boundaries as a femur contour, since the detected femur boundaries are very close. The femur length is calculated as the distance between the two end points of the femur.

3. Experimental materials and results

3.1. Materials

The data were retrospectively collected from Lincoln County Hospital (ULH), United Lincolnshire Hospitals NHS Trust, with patient identification removed. Two datasets with a total of 60 fetal US images are used in our experiment, which include different gestational ages (20, 21, 28, 34, 35 weeks). For dataset A, there are 20 images of fetal heads and 20 images of fetal femurs. Each ultrasound image was graded as 'poor' or 'good' image quality by an expert; while dataset B includes another 20 images of fetal head and femur, among which there are 10 fetal head images and 10 femur images.

The US images are produced in clinical practice by trained sonographers using Toshiba Aplio 780, Toshiba Aplio 790, Toshiba Aplio 500 (Toshiba Medical Systems, Tokyo, Japan) and GE Voluson 730 (GE Healthcare, Pollards Wood, UK) Ultrasound machines. Images collected from Toshiba manufacturers are originally stored as DICOM format with image size of 717×538 pixels, while the images from GE manufacturers are stored as the same DICOM format with size of 800×600 pixels. In our experimental, to keep image size consistency, we cropped the GE image size to be the same size as the Toshiba image with 717×538 pixels.

For dataset A, the ground truths are provided by two experts, one senior (Expert 1) and one junior

(Expert 2). Both experts manually delineated the objects of interest (here, the objects are the fetal skull and femur) twice for each image. For fetal head US images, two experts manually fitted ellipses to the fetal head and then the BPD, OFD and HC were derived from the fitted ellipse. For fetal femur US images, the experts manually drew outlines of fetal femur for each image, the FL was obtained from the delineated femur. For dataset B, 20 images were recently collected from the hospital. One senior expert performed manual delineations for all the head and femur images in dataset B, which are used as ground truths. The whole dataset B is used for the purpose of independent evaluation of the method.

Two experiments are reported in this paper. In the first experiment, dataset A is used for training and validation of the algorithm. For each anatomic structure (head, femur), the dataset A is split into half for the purpose of training and testing. In the second experiment, the method was further validated on an independent data set (dataset B), in which manual segmentations and measurements were carried out by expert1.

3.2. Evaluation methods

In order to validate the performance of the automatic segmentation as well as measurements using our proposed method, we exploited three different criteria. First two criteria are region based measures (Udupa et al. 2006) and contour distance based measures (Heimann et al. 2009), both criteria were chosen to assess the performance of the proposed segmentation method. The Bland-Altman plots (Bland and Altman 1986) were selected as third criterion to assess the measurement agreement between the proposed method and experts. This method also can be used to assess the inter- and intra-observer difference.

The region based measure is to compare the region generated by automatic segmentation and manual segmentation. Several region based measure parameters (precision, accuracy, sensitivity and specificity) are used. The results per image are averaged to obtain the overall automatic segmentation performance compared with manual segmentation.

The distance based measure is based on the calculation of the difference in millimeters between manually and automatically segmented contours of structures. Several standard measures: the maximum symmetric contour distance (MSD), the average symmetric contour distance (ASD) and the Root Mean Square Symmetric Contour Distance (RMSD) are used. The MSD returns the maximum distance between two contours, a higher MSD value indicates larger differences between the contours. Therefore the MSD equals 0 is the best case scenario which means the two contours are identical. The parameter ASD is the average distance between the two contours. ASD equals to 0 if the two contours are identical. The RMSD is the variant of the ASD, the larger difference compared to the ground truth can be emphasized by a larger value of RMSD.

Additionally, we further compared the measurements derived from automatic segmentation results

to the clinical measurement produced by the different experts using Bland-Altman plots (Bland and Altman 1986) to find if the differences in measurements between machines and experts are significant.

3.3. Comparative experimental results

The proposed fetal head/femur segmentation and measurement method is evaluated on the dataset described in the previous section. Table 1 shows the average inter- and intra-observe difference of manual segmentation over all images for both fetal head and femur. The intra-observer difference presented in Table 1 reflects the different levels of the two experts' experiences, but measurement values related to the intra-observer difference are still close for the two experts. For fetal head segmentation, the overall precisions of two experts are over 97% with standard deviation of around 1%, with all the other parameters (i.e. accuracy, sensitivity and specificity) values reaching about 99%. In addition, all distance based measurements (i.e. MSD, ASD and RMSD) showing minor differences. For the fetal femur, the intra-observer difference of expert 1 is better than the expert 2 in terms of precision, accuracy, specificity, ASD and RMSD, along with their corresponding standard deviations. The inter-observer difference shown in the last column of table 1 also reflects similarities between two experts' segmentation performances (note that the manual segmentations of the head are significantly better than the femur in term of precision). These results will be used as a reference to comparatively assess the performance of automatic segmentation method.

Table 1. Intra- and inter-observer difference of manual segmentation for fetal head and femur

Parameters (Head)	Intra-observer difference -E1 ^a	Intra-observer difference- E2 ^a	Inter-observer difference- E1 vs E2
Pre.(%)	97.51% (1.19) ^b	97.29% (0.98)	95.00% (2.13)
Acc.(%)	99.62% (0.18)	99.57% (0.15)	99.21% (0.39)
Sen.(%)	99.09% (0.84)	98.93% (1.06)	99.25% (0.92)
Spec.(%)	99.72% (0.22)	99.69% (0.13)	99.20% (0.53)
MSD(mm)	1.23 (0.62)	1.19 (0.39)	1.87 (0.89)
ASD(mm)	0.41 (0.20)	0.45 (0.16)	0.85 (0.40)
RMSD(mm)	0.52 (0.25)	0.55 (0.19)	0.98 (0.44)
(Femur)			
Pre.(%)	87.67% (2.33)	84.61% (3.8)	81.33% (5.4)
Acc.(%)	99.88% (0.02)	99.83% (0.06)	99.80% (0.08)
Sen.(%)	93.74% (2.56)	94.43% (2.8)	92.63% (2.6)
Spec.(%)	99.94% (0.02)	99.88% (0.07)	99.87% (0.07)
MSD(mm)	1.29 (0.49)	1.30 (0.28)	2.38 (1.19)
ASD(mm)	0.24 (0.06)	0.33 (0.10)	0.40 (0.15)
RMSD(mm)	0.34 (0.07)	0.43 (0.12)	0.58 (0.23)

^a E1 = Expert 1, E2 =Expert 2.

^b Results are given as the average value (standard deviation) for each measure

The evaluation results of the proposed segmentation method for fetal heads and femurs are summarized in Table 2. Given the intra observer difference of each expert and the inter observer

difference between two experts, two manual segmentation results which were delineated twice by each expert are defined as two ground truths, and then we compared our automatic segmentation results to each ground truth separately. There are four sets of evaluation results shown in each column of the table 2. The results show that on average, the fetal head segmentation performance of our method using the expert 1 as ground truth is closer to the inter-observer difference presented in table 1. However, it is also noted that, when we use expert 2 as ground truths, the head segmentation performance is much better than the inter-observer difference and is close to the intra-observer difference, where MSD, ASD and RMSD (for table 2 GT=E2 1st and GT=E2 2nd) are similar to the ones shown in table 1 (for both expert 1 and expert 2 intra-observer difference). This demonstrates a good performance of our method for fetal head segmentation. Additionally, when our method is compared to expert 1 (our method vs expert 1), a very low standard deviation (on average, about 1 mm for MSD, ≤ 0.5 mm for ASD and about 0.6 mm of RMSD) can be achieved. This demonstrates the robustness of our proposed method. For the evaluation of fetal femur segmentation, we use the same approach as that for the fetal head. The evaluation results obtained by comparing our method to expert 1 are closer to the intra-observer difference of expert 2; meanwhile the results are significantly better than the inter-observer difference between two experts. Note that if we define the expert 1 as ground truth, most of the standard deviations (i.e. precision, accuracy, MSD, ASD) in Table 2 femur section (in column 1 or column 2) are smaller than the corresponding values presented in last column of Table 1. This also reflects the stability of our method for fetal femur segmentation.

Table 2 Region based and distance based measures of our method for the fetal head and femur

Parameters	(GT= E1 1 st) ^a	(GT=E1 2 nd)	(GT=E2 1 st)	(GT=E2 2 nd)
(Head)				
Pre. (%)	94.91% (2.57)	94.38% (1.94)	96.85% (1.37)	96.30% (1.06)
Acc.(%)	99.19% (0.47)	99.12% (0.34)	99.50% (0.23)	99.41% (0.17)
Sen.(%)	99.16% (1.18)	99.23% (1.27)	98.06% (1.37)	98.08% (1.41)
Spec.(%)	99.19% (0.62)	99.09% (0.45)	99.76% (0.16)	99.65% (0.19)
MSD(mm)	2.37 (1.46)	2.58 (0.97)	1.46 (0.56)	1.63 (0.51)
ASD(mm)	0.87 (0.48)	0.94 (0.35)	0.53 (0.24)	0.62 (0.19)
RMSD(mm)	1.08 (0.64)	1.18 (0.44)	0.67 (0.28)	0.76 (0.25)
(Femur)				
Pre. (%)	84.37% (3.1)	83.91% (3.5)	74.97% (7.02)	77.50% (5.0)
Acc.(%)	99.85% (0.04)	99.85% (0.04)	99.73% (0.1)	99.77% (0.08)
Sen.(%)	88.68% (3.88)	88.16% (3.96)	78.68% (8.7)	82.18% (6.4)
Spec.(%)	99.96% (0.02)	99.96% (0.02)	99.95% (0.02)	99.95% (0.03)
MSD(mm)	2.72 (0.99)	2.68 (0.99)	3.29 (1.10)	3.10 (1.35)
ASD(mm)	0.31 (0.07)	0.32 (0.08)	0.56 (0.21)	0.48 (0.16)
RMSD(mm)	0.52 (0.14)	0.51 (0.15)	0.81 (0.32)	0.72 (0.25)

^a GT=E1 1st denotes that the manual segmentations provided by expert 1 at the first time are used as ground truth (GT).

Moreover, the intra- and inter- observer differences of the fetal head biometric measurements were assessed using Bland-Altman plots. Examples of these plots are illustrated in Figure 9, where (a)

denotes the inter-observer difference of the BPD for the two experts, and (b) is the measurement difference between the expert 1 and our automated method.

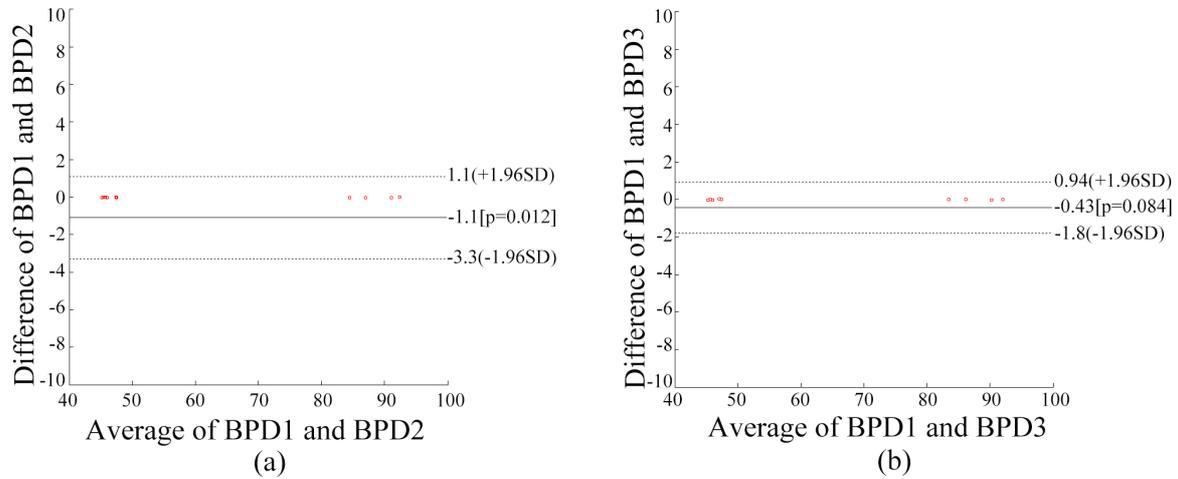


Figure 9 Example of the Bland-Altman plots for BPD. (a) the inter-observer difference between expert 1 and expert 2, where BPD1 denotes the measurement from expert 1, while BPD 2 from expert 2. (b) the difference between expert 1 and our method, where BPD 3 is from our automated method.

For two manual measurements (BPD1 and BPD2), the mean difference of BPD is -1.1 mm with 95% confidence interval from 1.1 mm to -3.3 mm, at $p=0.012$ confident level, as shown in Figure 9 (a). While the 95% limits of agreements for inter observer difference between automated and expert 1 measurements for BPD are from 0.94 to -1.8 mm, with -0.43 of the mean difference, at $p=0.084$ confident level, as shown in Figure 9(b). It is noted that, comparing Figure 9 (a) with (b), the smaller mean difference and the relatively narrow limits shown in (b) indicate a good level of agreement between the automated and expert 1’s measurements.

For comparison, the intra- and inter-observer difference of manual biometric measurements for fetal head and femur are shown in Table 3. The evaluation results of our method compared to two experts’ ground truths are shown in Table 4.

For the fetal head clinical measurements, comparing with two separate delineations for each expert, the differences of BPD, OFD and HC are very close, at < 1 mm level; however there are minor differences in these measures between the two experts. In Table 4, we observe that, the BPD measures produced by our method are closer to the two experts’ measures, while using the expert 2’s measures as ground truth, overall our method is better than the ones using the expert 1’s measures. It can also be noted from Table 4 that the manual segmentation of the skulls produced by expert 1 are smaller than the automated segmentation. However, the major differences of HC are caused by the differences of OFD, the differences of BPD maintains at < 1 mm level (refer to the first row in table 4). This can be seen in the segmentation examples shown in the Fig. 8. A further discussion is given in the next Section.

Table 3 Intra- and inter-observer difference of manual biometric measurements for fetal head and femur

Parameters (Head)	Intra-observer difference- E1 ^a	Intra-observer difference- E2	Inter-observer difference- E1 vs E2
BPD(mm)	-0.20 (0.41) ^b	0.31 (0.99)	-1.10 (1.11)
OFD(mm)	-0.34 (1.50)	-0.60 (0.52)	-1.86 (2.42)
HC(mm)	-0.83 (2.77)	0.62 (1.71)	-4.68 (5.45)
(Femur)			
FL(mm)	0.40 (0.54)	0.14 (0.66)	0.44 (1.33)

^a E1 = Expert 1, E2 =Expert 2.

^b Results are given as the average value (standard deviation) for each measure

For the fetal femur measure (FL), we can find in table 4 that the differences between our method and experts, on average, are at <1 mm level. While comparing our method to expert 1, the corresponding standard deviations (1.47, 1.53) are very close to the inter-observer difference.

Table 4 Evaluation results of our method for fetal head and femur biometric measurements

Parameters (Head)	(GT=E1 1 st) ^a	(GT=E1 2 nd)	(GT=E2 1 st)	(GT=E2 2 nd)
BPD(mm)	-0.43 (0.70)	-0.63 (0.71)	-0.36 (0.75)	-0.67 (0.84)
OFD(mm)	-2.57 (3.71)	-2.91 (3.15)	0.10 (1.67)	0.71 (1.86)
HC(mm)	-4.54 (6.67)	-5.37 (5.57)	-0.76 (2.44)	-0.14 (3.13)
(Femur)				
FL(mm)	-0.19 (1.47)	-0.59 (1.53)	0.49 (1.72)	0.63 (1.62)

^a GT=E1 1st denotes that the manual segmentations provided by expert 1 at the first time are used as ground truth (GT).

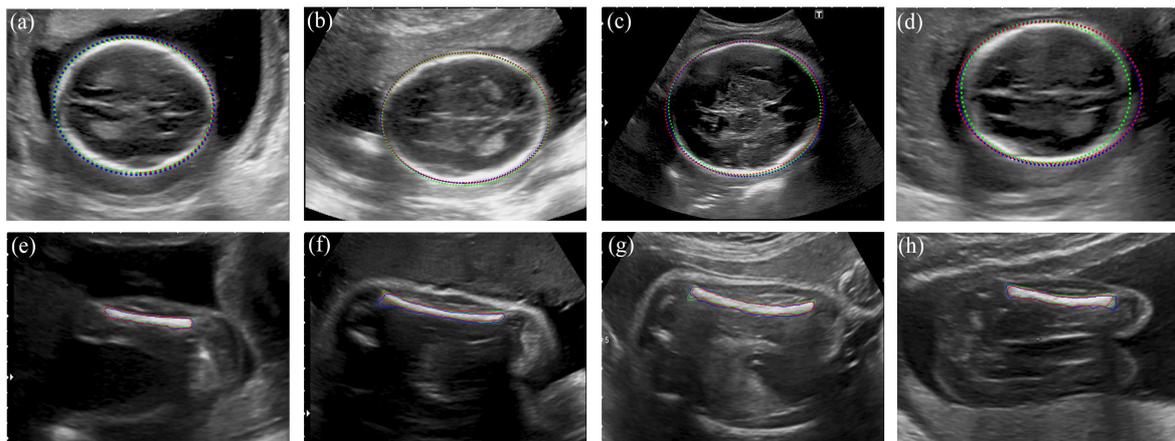


Figure 10. The best and worst cases of our automatic segmentations over two experts' segmentations. The contours of structures produced by the automatic method is shown in red, contours delineated by expert 1 and 2 are shown in green and blue, respectively. (a)(b) are the best cases of fetal head segmentations, (c)(d) are the worst cases, (e)(f) are the best cases for fetal femur segmentation, (g)(h) are the worst femur segmentation cases.

The top row of Fig. 10 illustrates the best and worst fetal head segmentation cases in our experiments, as we can see that three head contours of best cases in Fig. 10 (a) (b) which are produced by the two experts and our method are almost identical with very tiny differences. A few clear differences in the horizontal distances can be observed in the worst cases (Fig. 10 (c)(d)). The best and worst fetal femur segmentation cases are illustrated in the bottom row of Fig. 10. As we can see from Fig. 10 (e)(f), the three contours are very close to each other, the contours in Fig. 10 (g)(h) show minor differences between our method and both experts.

Additionally, our method was further tested on another independent dataset (dataset B). The evaluation results for segmentations and measurements are summarized in tables 5 and 6, respectively. The results were evaluated by comparing to the ground truths produced by expert 1. Without re-training the model, the overall performance of our method for this independent dataset is a little worse than the previous results (refer to the tables 2 and 5) as expected, but the average precisions of head and femur segmentation reach 91.50% and 77.63%, respectively, whilst the values of MSD, ASD and RMSD are close to the values obtained using training manner which are presented in the Table 2. These evaluation results demonstrate the robustness of our method.

Table 5 Region based and distance based measures of our method for the fetal head and femur using the independent data set (dataset B).

Parameters	Our method (for Head)	Our method (for Femur)
Pre. (%)	91.50% (2.91)	77.63% (1.29)
Acc. (%)	98.88% (0.36)	99.80% (0.07)
Sen. (%)	95.82% (3.76)	95.19% (3.45)
Spec. (%)	99.33% (0.42)	99.84% (0.08)
MSD(mm)	3.30 (1.09)	2.00 (0.76)
ASD(mm)	1.34 (0.44)	0.51 (0.52)
RMSD(mm)	1.64 (0.55)	0.61 (0.37)

Table 6 Evaluation results of our method for fetal head and femur biometric measurements using the independent data set (dataset B).

Parameters (for Head)	Our method
BPD(mm)	0.26 (2.46)
OFD(mm)	-0.65 (4.61)
HC(mm)	-0.22 (9.53)
Parameters (for Femur)	
FL(mm)	-0.67 (1.06)

4. DISCUSSION AND CONCLUSION

In this paper, we present an automatic segmentation method for both head and femur measurements in fetal US images. The method is validated against manual segmentations provided by two experts. In particular, region-based and contour-based evaluation methods are used to validate the performance of our segmentation method. Meanwhile, the biometric measurements (for head: BPD,

OFD, HC, and for femur: FL) derived from the segmentations are also compared to manual analysis. The experimental results have shown much promise for clinical applications.

Rueda et al. (2014) organized a grand challenge US –ISBI 2012. The challenge was closed and the dataset used in the challenge is not publicly available. Therefore, it is difficult to provide a direct comparison to other methods using that particular dataset. However, it is noted that, although the datasets are different, the similarities of intra- and inter-observer difference of manual segmentations between our and the challenge datasets indicate the close similarity of the two datasets. The image size (717×538) in our dataset is also similar to the image size (756×546) used in their work (Rueda et al. 2014). Therefore, the evaluation results as well as the method efficiencies presented in the grand challenge (Rueda et al. 2014) can provide a comparable reference. For the fetal head, the segmentation performance of our method is close to the best method proposed by the Foi et al. (2014) in ISBI 2012, in which they achieved 95.72%(1.92) precision, 2.16(1.44) MSD, and 0.88(0.53) ASD. For fetal femur segmentation, our evaluation results are better than the results of all proposed femur segmentation methods in the challenge (Rueda et al. 2014). The best reported femur segmentation method achieved 65.44% (16.98) precision, 6.39 (9.53) MSD, and 1.23(2.3) ASD. Our algorithm was implemented using Matlab 2011b. The current computation time was on an average of 163s per image on a laptop (Mackbook pro, 2.53, Inter Corse i5, Mac os x). This is accumulated from each step discussed in section 2, where the pre-processing step (speckle noise reduction) is 6.4s, the total execution time for feature extraction, texton map generation and boundary detection is 152.5s, while boundary identification and fitting steps take 4.8s. Comparing to the other methods on the efficiency reported in (Rueda et al. 2014), our current efficiency is on average, which is better than 196s of Ciurte et al. (2012a) and 5 min of Foi *et al* (2014), and closer to 100-150s of the Stebbing and McManigle (2012). It is noted that the accuracy and the reproducibility of the method is the main aim of this work. In the next step, we will optimize the current code, this could considerably reduce the computation time.

We should point out, the evaluation results of our method (Table 2) compared to expert 2 measurements are much better than the those using expert 1 measurement for fetal head segmentation, this is due to expert 1 adhered more closely to the two orthogonal diameters, the BPD and OFD, so their ground truth ellipse fitted closely to the skull at the points of BPD and OFD measurement. The ellipse of expert 2 was a more holistic fit to the entire shape of the skull. While, our proposed method also looked for the entire shape of the skull, i.e. segmenting the initial boundaries of fetal structures, and then using an ellipse fitting method to construct a closed head contour. This is also one of the main reasons that overall the manual segmentation of the skulls produced by expert 1 are smaller than the automated segmentation (Table 4). In addition, for fetal head biometric measurements in our experiment, the differences in HC are primarily caused by the differences in the OFD, which are derived from the disagreements of diagnosis of bone structures in the areas where skulls have fuzzy

boundaries. Commonly, these phenomena are characterized as flocculent structures surrounding the highlighted bone structures in which the discriminations between bone and non-bone structures cannot be pre-attentively perceived. Moreover, normally, the skull presented in a US image is an incomplete structure in the areas around the horizontal left and right points, while the ellipse is fitted (automatically or manually) based on features of visible structures in US images (here these structures are top and bottom parts of skull), this may lead to disagreements in the ellipse estimations among experts. The improvement of automatic fetal head segmentation problem can be linked to the definition of a specific criterion in clinical applications for distinguishing between bones and non-bones structures in fetal US images. Alternatively, an image quality assessment guideline could be established to assist the operator in producing better quality fetal US images, such work has been reported in (Dudley and Chapman 2002).

We should also mention that, in current obstetrics workflow, statistically significant difference between the measurements by two different methods does not imply that the difference would affect clinical decisions. Likewise, even if the measurement differences were not statistically significant, they could still result in different clinical decisions depending on the patient and the nature of the decision being made (Foi et al. 2014). The current GA estimation depends on several factors, including fetal biometric measurements, the regression equations and charts used to estimate GA. In the long term, with the improvement of the methods used to calculate GA, more accurate and reproducible measurement would provide more accurate GA of the fetus.

The practical value of this automatic system is to remove the dependence of measurements on operator judgement and track-ball sensitivity, providing more accurate and reproducible results. There will also be a timesaving, allowing sonographers to focus more on image acquisition and interpretation. Our future work will include code optimization to improve the efficiency, automated US image quality assessment and further evaluation of the method in a large clinical study.

Acknowledgment

This work was supported by the United Lincolnshire Hospitals NHS Trust. Thanks to the sonographers of United Lincolnshire Hospitals NHS Trust, for providing the US scans for this project.

References

- Arbelaez P, Maire M, Fowlkes C, Malik J. Contour detection and hierarchical image segmentation. *IEEE Trans Pattern Anal Mach Intell* 2011;33:898-916.
- Bland JM, Altman DG. Statistical Methods for Assessing Agreement between Two Methods of Clinical Measurement. *Lancet* 1986;1:307-10.
- Carneiro G, Georgescu B, Good S. Knowledge-based Automated Fetal Biometrics. Whitepaper ACUSON S2000TM Ultrasound System 2008a;1-5.
- Carneiro G, Georgescu B, Good S, Comaniciu D. Detection and measurement of fetal anatomies from ultrasound images using a constrained probabilistic boosting tree. *IEEE transactions on medical imaging* 2008b;27:1342-55.
- Chaudhuri S, Chatterjee S, Katz N, Nelson M, Goldbaum M. Detection of blood vessels in retinal images using two-dimensional matched filters. *IEEE transactions on medical imaging* 1989;8:263-9.
- Ciurte A, Bresson X, Cuadra MB. A semi-supervised patchbased approach for segmentation of fetal ultrasound imaging. *Proceedings of Challenge US: Biometric Measurements from Fetal Ultrasound Images, ISBI 2012* 2012a;5-7.
- Ciurte A, Rueda S, Bresson X, Nedeveschi S, Papageorghiou AT, Noble JA, Bach Cuadra M. Ultrasound Image Segmentation of the Fetal Abdomen: a Semi-Supervised Patch-Based Approach. *Proc the MICCAI Workshop on PaPI 2012* 2012b;1-8.
- Dudley NJ, Chapman E. The importance of quality management in fetal measurement. *Ultrasound Obst Gyn* 2002;19:190-6.
- Espinoza J, Good S, Russell E, Lee W. Does the Use of Automated Fetal Biometry Improve Clinical Work Flow Efficiency? *J Ultras Med* 2013;32:847-50.
- Fitzgibbon AW, Pilu M, B. FR. Direct Least Squares Fitting of Ellipses. *IEEE Trans on Pattern Analysis and Machine Intelligence* 1999;21:476-80,.
- Foi A, Maggioni M, Pepe A, Rueda S, Noble JA, Papageorghiou AT, Tohka J. Difference of Gaussians revolved along elliptical paths for ultrasound fetal head segmentation. *Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society* 2014;38:774-84.
- Haralick RM, Shanmuga K, Dinstein I. Textural Features for Image Classification. *IEEE T Syst Man Cyb* 1973;Smc3:610-21.
- Heimann T, van Ginneken B, Styner MA, Arzhaeva Y, Aurich V, Bauer C, Beck A, Becker C, Beichel R, Bekes G, Bello F, Binnig G, Bischof H, Bornik A, Cashman PMM, Chi Y, Cordova A, Dawant BM, Fidrich M, Furst JD, Furukawa D, Grenacher L, Hornegger J, Kainmuller D, Kitney RI, Kobatake H, Lamecker H, Lange T, Lee J, Lennon B, Li R, Li S, Meinzer HP, Nemeth G, Raicu DS, Rau AM, van Rikxoort EM, Rousson M, Rusko L, Saddi KA, Schmidt G, Seghers D, Shimizu A, Slagmolen P, Sorantin E, Soza G, Susomboon R, Waite JM, Wimmer A, Wolf I. Comparison and Evaluation of Methods for Liver Segmentation From CT Datasets. *IEEE transactions on medical imaging* 2009;28:1251-65.

Jardim SM, Figueiredo MA. Segmentation of fetal ultrasound images. *Ultrasound in medicine & biology* 2005;31:243-50.

Julesz B. Textons, the elements of texture perception, and their interactions. *Nature* 1981;290:91-7.

Leung T, Malik J. Representing and recognizing the visual appearance of materials using three-dimensional textons. *Int J Comput Vision* 2001;43:29-44.

Loughna P, Chitty L, Evans T, Chudleigh T. Fetal size and dating: Charts recommended for clinical obstetric practice. *Ultrasound* 2009;17:160-6.

Lu W, Tan JL, Floyd R. Automated fetal head detection and measurement in ultrasound images by iterative randomized Hough transform. *Ultrasound in Medicine and Biology* 2005;31:929-36.

Malik J, Belingie S, Shi J, Leung T. Textons, Contours and Regions: Cue Integration in Image Segmentation. *Proc IEEE Intl Conf Computer Vision* 1999;918-25.

Martin DR, Fowlkes CC, Malik J. Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE Trans Pattern Anal Mach Intell* 2004;26:530-49.

Namburete AIL, Noble JA. Fetal Cranial Segmentation in 2d Ultrasound Images Using Shape Properties of Pixel Clusters. *I S Biomed Imaging* 2013;720-3.

Pemberton LK, Burd I, Wang E. An appraisal of ultrasound fetal biometry in the first trimester. *Reports in Medical imaging* 2010;2010:11-5.

Perona P, Malik J. Scale-Space and Edge-Detection Using Anisotropic Diffusion. *IEEE T Pattern Anal* 1990;12:629-39.

Petkov N. Biologically Motivated Computationally Intensive Approaches to Image Pattern-Recognition. *Future Gener Comp Sy* 1995;11:451-65.

Petkov N, Kruijinga P. Computational models of visual neurons specialised in the detection of periodic and aperiodic oriented visual stimuli: bar and grating cells. *Biological cybernetics* 1997;76:83-96.

Rueda S, Fathima S, Knight CL, Yaqub M, Papageorghiou AT, Rahmatullah B, Foi A, Maggioni M, Pepe A, Tohka J, Stebbing RV, McManigle JE, Ciurte A, Bresson X, Cuadra MB, Sun CM, Ponomarev GV, Gelfand MS, Kazanov MD, Wang CW, Chen HC, Peng CW, Hung CM, Noble JA. Evaluation and Comparison of Current Fetal Ultrasound Image Segmentation Methods for Biometric Measurements: A Grand Challenge. *IEEE transactions on medical imaging* 2014;33:797-813.

Shrimali V, Anand RS, Kumar V. Improved segmentation of ultrasound images for fetal biometry, using morphological operators. *Conference proceedings : Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Annual Conference* 2009;2009:459-62.

Stebbing RV, McManigle JE. A boundary fragment model for head segmentation in fetal ultrasound. *Proceedings of Challenge US: Biometric Measurements from Fetal Ultrasound Images, ISBI 2012* 2012;9-11.

Udupa JK, LeBlanc VR, Ying ZG, Imielinska C, Schmidt H, Currie LM, Hirsch BE, Woodburn J. A framework for evaluating image segmentation algorithms. *Comput Med Imag Grap* 2006;30:75-87.

Varma M, Zisserman A. Unifying statistical texture classification frameworks. *Image Vision Comput* 2004;22:1175-83.

Weickert J. Recursive separable schemes for nonlinear diffusion filters. *Lect Notes Comput Sc* 1997a;1252:260-71.

Weickert J. A review of nonlinear diffusion filtering. *Lect Notes Comput Sc* 1997b;1252:3-28.

Yu JH, Wang YY, Chen P. Fetal ultrasound image segmentation system and its use in fetal weight estimation. *Med Biol Eng Comput* 2008;46:1227-37.

Zhang F, Yoo YM, Koh LM, Kim Y. Nonlinear diffusion in Laplacian pyramid domain for ultrasonic speckle reduction. *IEEE transactions on medical imaging* 2007;26:200-11.

Zhu SC, Guo CE, Wang YZ, Xu ZJ. What are textons? *Int J Comput Vision* 2005;62:121-43.