DIFFERENCE OF GAUSSIANS REVOLVED ALONG ELLIPTICAL PATHS FOR ULTRASOUND FETAL HEAD SEGMENTATION

Alessandro Foi[a], Matteo Maggioni[a], Antonietta Pepe[a], Sylvia Rueda[b], J. Alison Noble[b], Aris T. Papageorghiou[c], Jussi Tohka[a]

[a]*Department of Signal Processing, Tampere University of Technology, P.O. Box 553, FIN-33101, Finland. Corresponding author: Jussi Tohka Email address: jussi.tohka@tut.fi, fax: +358-3-31153087, tel.+358-40-1981497*
[b]*Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford, U.K.*
[c]*Nuffield Department of Obstetrics and Gynaecology and Oxford Maternal and Perinatal Health Institute, Green Templeton College, University of Oxford, Oxford, U.K.*

Abstract

We present a fully automatic method to segment the skull from 2-D ultrasound images of the fetal head and to compute the standard biometric measurements derived from the segmented images. The method is based on the minimization of a novel cost function. The cost function is formulated assuming that the fetal skull has an approximately elliptical shape in the image and that pixel values within the skull are on average higher than in surrounding tissues. The main idea is to construct a template image of the fetal skull parametrized by the ellipse parameters and the calvarial thickness. The cost function evaluates the match between the template image and the observed ultrasound image. The optimum solution that minimizes the cost is found by using a global multiscale, multistart Nelder-Mead algorithm. The method was qualitatively and quantitatively evaluated using 90 ultrasound images from a recent segmentation grand challenge. These images have been manually analyzed by 3 independent experts. The segmentation accuracy of the automatic method was similar to the inter-expert segmentation variability. The automatically derived biometric measurements were as accurate as the manual measurements. Moreover, the segmentation accuracy of the presented method was superior to the accuracy of the other automatic methods that have previously been evaluated using the same data.

Key words: biparietal diameter; head circumference; energy minimization; fetal biometry; image analysis; global optimization;

1. Introduction

Ultrasound measurements of fetal biometry is a standard method for dating pregnancies and for assessment of fetal growth. Standard biometric measurements include biparietal diameter (BPD), occipito-frontal diameter (OFD), head-circumference (HC), abdominal circumference (AC), crown-rump length

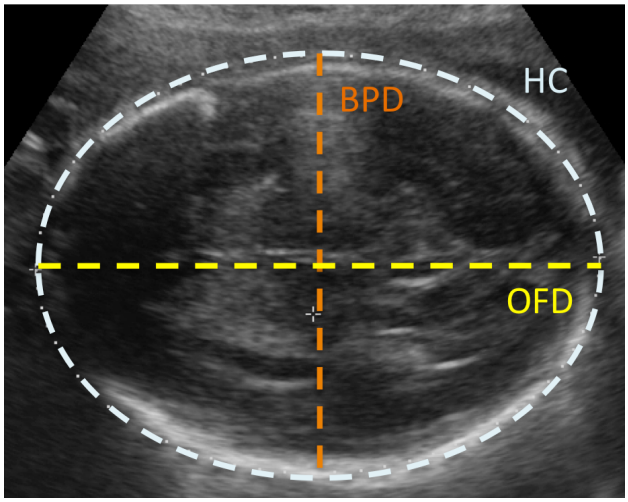


Figure 1: The standard biometric measurements of the fetal head on a 2-D ultrasound image. The biparietal diameter (BPD) is measured as the distance between the outer borders of the parietal bones (outer to outer) at the widest part of the skull. The occipitofrontal distance (OFD) is measured between the outer borders of the occipital and frontal edges of the skull at the point of the midline (outer to outer) across the longest part of the skull. HC is computed based on OFD and BPD as $\text{HC} = \pi(\text{OFD} + \text{BPD})/2$.

(CRL) and femur length (FL). Based on these measurements, the gestational age and size of the fetus can then be estimated using charts or regression formulae [5, 15, 8]. In particular, biometrics related to the fetal head, i.e., BPD and HC as shown in Figure 1, are recommended for measuring the gestational age during second or third semester and are used for assessing the fetal size [5, 15, 13].

Currently, expert users perform these measurements manually. This is not only time consuming but also it results in high intra- and inter-observer variability of these measurements. Thus, automating measurement processes could provide significant benefits to both pregnant women and clinicians [2]. However, designing a fully automatic procedure is a challenging task because of the highly variable image quality that is a typical characteristic of ultrasound images. In addition, various image artefacts and surrounding tissues further complicate the automatic measurement of fetal biometrics. These problems are easy to appreciate in Figure 2, whose panels (a) and (b) show two low quality but still typical ultrasound images.

Given the importance of the problem, several automatic and semi-automatic methods for computing biometric measurements from ultrasound images have been proposed. These methods have mostly been based on the segmentation of the ultrasound images [22]. Approaches for the fetal skull segmentation have been based on deformable spline [12] and contour models [4, 20], supervised learning [2], and ellipse fitting through (randomized) Hough transform [10, 17, 27]. The deformable contour or spline

model [4, 20, 12] based approaches rely on the user to indicate a point close to the center of head and construct an initial contour model based on this information. This initial contour model is then deformed by minimizing a cost function.

The ellipse fitting approaches first segment pre-processed ultrasound images through clustering or histogram analysis [10, 17] or perform edge detection [27], in either case producing a binary image of the fetal skull. The Hough transform [10] or the randomized Hough transform (RHT) [17, 27] is then utilized to fit the ellipse model to the binary image. As Hough transform and RHT are very sensitive to image artifacts, these methods require the user to define an initial region of interest (ROI) [27] or their application is limited to images where head contour is complete and is not overlapped by other structures [10].

The supervised learning approach of [2] is different from the methods cited above as it directly exploits expert annotations of the images to produce the segmentations. More specifically, the method of [2] searches for a minimal bounding box containing the head of the fetus in the ultrasound image. The method works by training a three-stage classifier, termed constrained probabilistic boosting tree, based on manually segmented images.

In this paper, we present a new method for the fetal skull segmentation from 2-D ultrasound images. Based on the skull segmentations, it is then straightforward to compute the biometric measurements of BPD, OFD, and HC. The method is completely automatic and it requires no user interaction unlike many methods cited above. This new method is named Difference of Gaussians revolved along Elliptical paths or DoGEll. The idea is to construct a template image of the fetal skull parametrized by both the calvarial thickness (the thickness of the skull) and an ellipse modelling the contour of the skull. This is unlike previous ellipse fitting methods [10, 17, 27] that try to fit a single ellipse into the image, i.e., these previous methods do not model the skull as an entity having a finite thickness but instead make use of image processing operations that try to reduce its thickness to a single pixel. In constructing the template image, we assume that the image intensity is high within pixels representing the skull (due to a high acoustic impedance of the bone) and lower immediately inside and outside the skull. The template is then matched to the ultrasound image by minimizing a cost function designed to measure the lack of match between our parametric elliptical image model and the observed image. More specifically, for a given ellipse, we construct a surface that models the pixel values of the skull and surrounding areas by revolving a difference of Gaussians (DoG) along the elliptical path and define the cost function as a negative correlation between the observed image and the surface. This cost function is robust to various image imperfections as demonstrated in Fig. 2. To find the optimum parameters defining the ellipse and calvarial thickness, the cost function is globally minimized by a multiscale multistart Nelder-Mead algorithm [19, 14] devised specifically for this purpose. Notice that this new method differs from the previous region based methods [12, 2] in that instead of trying to maximize the match the image segmentation (consisting of foreground and background regions) as in the previous methods, it matches a template image constructed based on the skull parameters to the observed image. This way it is possible to reduce the sensitivity of the method to the high intensity artifacts in the background region.

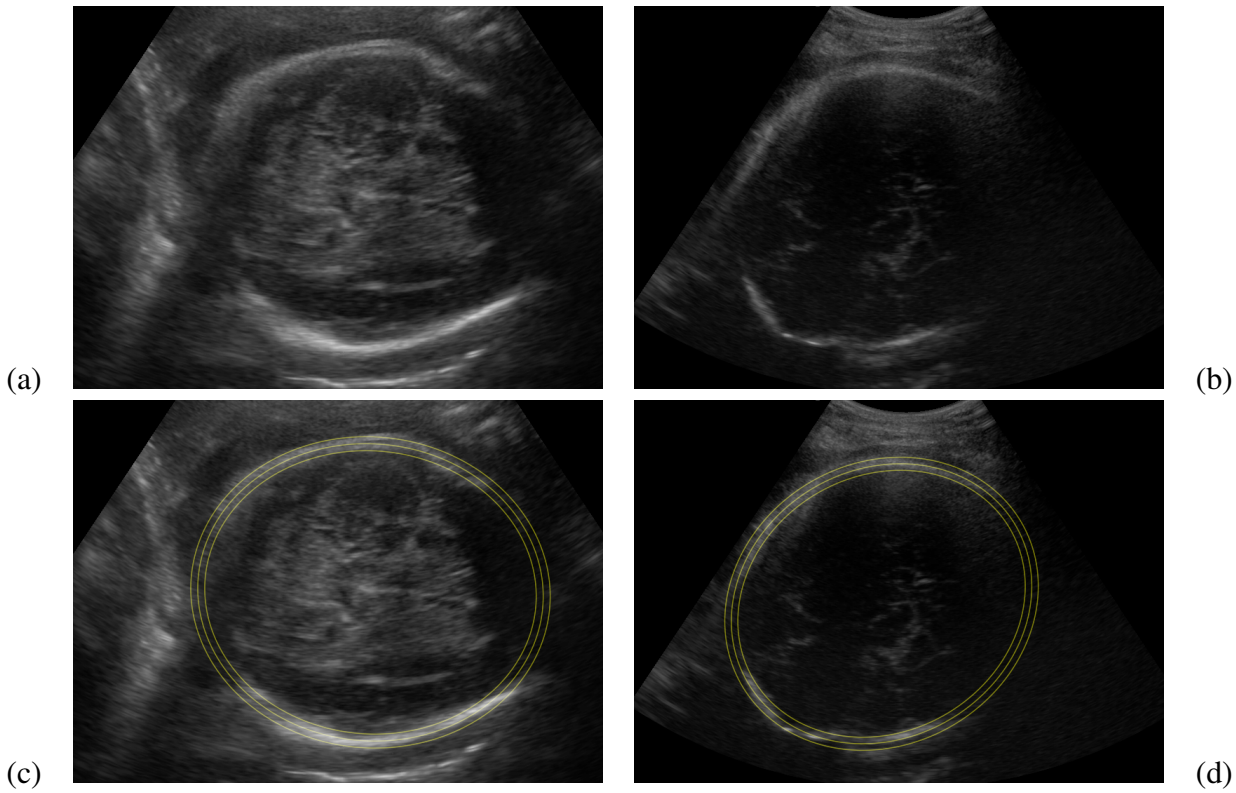


Figure 2: Two example images from the ISBI 2012 ultrasound segmentation challenge [22] and their DoGEll segmentations. (a) 28 week old fetus. (b) 33 week old fetus. In both panels (a) and (b), the skull is only partially visible, has different contrast in different image regions, and there are artifacts with bright intensity in the images that create a considerable challenge to automatic segmentation methods. (c) and (d): DoGEll segmentations of these two images demonstrating the robustness of DoGEll against image artifacts. The three yellow contours are the estimates of the inner, medial, outer contour of the skull produced by the DoGEll method.

This DoGEll method won the first prize of the head sub-challenge in the *Challenge US: Biometric Measurements from Fetal Ultrasound Images*, held in conjunction with and supported by the *IEEE International Symposium on Biomedical Imaging* (ISBI) 2012 in Barcelona, Spain [22]. The contributions of the present paper are two-fold. First, we present a detailed account of the winning DoGEll method for the first time. The method has not been published in peer-reviewed literature. A short account of it has been presented in the Challenge proceedings [9] and one paragraph summary was presented in [22]. Second, we present a new, improved implementation of the DoGEll method typically running in under 5 seconds per image, thus yielding a 60 fold savings in computation time as compared to the one winning the segmentation challenge. The difference in the quantitative results between the implementations due to speed ups is minimal and the new, faster implementation would have won the segmentation challenge as well. The main specific differences between the old [9] and new (this paper) method are i) the new minimization algorithm that uses considerably fewer restarts and embeds an insertion sort strategy to accelerate the Nelder-Mead algorithm, ii) a cost function (Eq. (3)) that is modified to be better principled mathematically as explained in more detail in Section 2, and iii) faster image pre-processing.

The evaluation of our method is carried out with the dataset used in the *Challenge US: Biometric Measurements from Fetal Ultrasound Images*. This consists of 90 2-D fetal ultrasound images of the head, some of which are purposely of very low quality, acquired on fetuses at different gestational ages. The segmentations and biometric parameters derived from the automatic DoGEll method are compared to the manual analysis of three independent experts. The evaluation in this paper is based on the same principles as in the challenge, i.e., the method development and evaluation were completely independent and the method developers (AF, MM, AP and JT) did not have access to the manual segmentations nor to the biometric measurements. The evaluation results show that the DoGEll method can automatically segment the fetal skull from ultrasound images with an accuracy similar to the inter-expert variability of the manual segmentation.

The paper is organized as follows. In Section 2, we describe the cost function that we minimize to obtain the segmentations. In Section 3, we describe the applied image pre-processing, the algorithm for the minimization of the cost function, and the procedure for computation of biometric parameters based on the segmentation results. Section 4 describes the quantitative validation of the method using ISBI 2012 Challenge data, Section 5 presents the results of the validation, and Section 6 concludes the paper by discussing the methodological contributions and experimental results.

2. Cost function

In this section, we describe the cost function to segment fetal skulls from ultrasound images. As already mentioned, we assume that the head contour of the fetus can be modeled by an ellipse. The main rationale behind the proposed cost function, as illustrated in Fig. 3, consists of fitting of an image model which comprises of three nested elliptical annuli to the image: the centermost representing the skull of the fetus characterized by high image intensity values (bright pixels); the inner and the outer

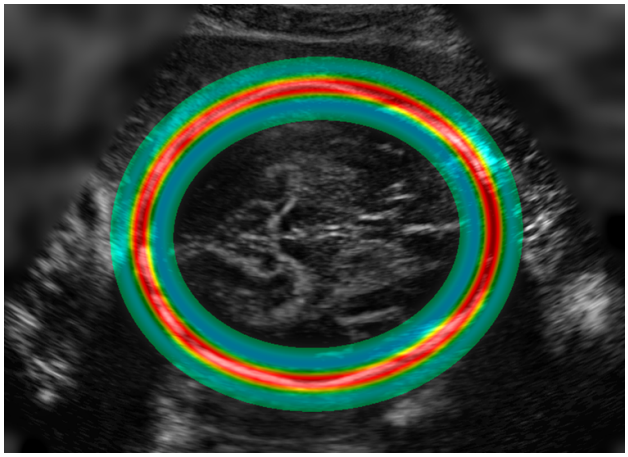


Figure 3: An illustration of the principal idea of the cost function. The red region in the figure models the fetal skull and we expect the pixel values in this area to be high. The green regions model the regions surrounding the skull where the pixel values are to expected to be lower than in the skull. The transition between the skull region (in red) and background (in green) is smooth (see also Fig 4.) and this border area is colored in yellow. The pixel values in the non-colored region contribute minimally to the value of cost function.

representing the surrounding areas, usually exhibiting relatively low image intensity values. The image model is constructed based on a single ellipse and a parameter modeling the thickness of the skull. We use a difference of Gaussians, see Fig. 4 (a), along each half line leaving from the center of ellipse in modelling the regions.

2.1. Image model

In more detail, we parametrize each ellipse $E(\mathbf{a})$ with 5 parameters: center coordinates c_1, c_2, semi-axis lengths r_1, r_2, and rotation angle in radians θ, organized into the vector

$$\mathbf{a} = [c_1, c_2, r_1, r_2, \theta].$$

Let $h_{(x_1,x_2,\mathbf{a})}$ be the radial half-line leaving from the center (c_1, c_2) and passing through the point (x_1, x_2). Using $h_{(x_1,x_2,\mathbf{a})}$, we measure the radial distance $d(x_1, x_2, \mathbf{a})$ between (x_1, x_2) and the ellipse $E(\mathbf{a})$, as well as the normalized radial distance $r_0(x_1, x_2, \mathbf{a})$ between (x_1, x_2) and (c_1, c_2), i.e. the distance between those two points divided by the radius of the ellipse along $h_{(x_1,x_2,\mathbf{a})}$. Then we define a surface

$$g(x_1, x_2, \mathbf{a}, s) = \frac{f_s(d(x_1, x_2, \mathbf{a})) - f_{3s}(d(x_1, x_2, \mathbf{a}))}{r_0(x_1, x_2, \mathbf{a})}, \tag{1}$$

where f_s and f_{3s} are two univariate Gaussians centered at zero with standard deviations equal to s and $3s$, respectively. This is our image model for a given ellipse. An example of the surface $g(x_1, x_2, \mathbf{a}, s)$ is shown in Fig. 4 (b) and the rationale for its design is described in Fig. 4 (c).

Before introducing the cost function based on the image model, we provide a few facts about the function g, including the choice of the standard-deviation parameters of the DoG, geometric properties of g, and its zero-mean radial property.

In DoGs, the ratio between the standard deviations of the two Gaussian functions mainly controls the rate at which the negative tails approach zero relative to the width of the central positive lobe. A larger ratio gives longer tails, and hence, in our case, more sensitivity to the surround distant from the skull. As the ratio approaches unity, the DoG approaches the minus second derivative of a Gaussian. In most applications the ratio takes values between 1.5 and 5. For our purposes a value of 3 provides sufficient sensitivity to the surrounding region of the skull [18, 16].

The numerator of g enjoys two distinct symmetries. First, it is radially symmetric, i.e., the profile along each radial line is one and the same: it is the DoG. Equivalently we can say that the numerator of g is constant along the contour lines. Second, each radial profile is symmetric about the point of intersection between the radius and the ellipse (a trivial consequence of the symmetry of the DoG). Overall, it means that the numerator of g assumes the same value on any two distinct contour lines sharing the same radial distance from the ellipse (one contour line internal to the ellipse, and the other one external). Even though the DoG is by construction a zero-mean function, after elliptical revolution the integral balance is lost due to the nonuniform radial geometry, as shown in Fig. 4(c)). In particular, the green patches illustrate two finite area elements. Riemann integration of a bivariate function is the limiting summation over these elements, as their respective area tends to zero. Observe in the figure that the size of the area elements increases proportionally to their radial distance from the center. Therefore, by dividing the numerator by r_0, we restore the integral balance.

Clearly, $\int_{-\infty}^{+\infty} f_s(\xi) - f_{3s}(\xi) d\xi = 0$. However, when defining g, each radial line is bounded, on one side by the center of the ellipse, and on the other side by the image boundary. This means that the argument $d(x_1, x_2, \mathbf{a})$ ranges from $d(c_1, c_2, \mathbf{a})$ to a finite positive value attained at the intersection of the radial line with the image boundary. Therefore, the tails of the DoG are formally missing from g. This notwithstanding, due to their exponential decay, the tails are numerically negligible for all parameter combinations of practical relevance[1]. Thus,

$$\iint_{\Theta} g(x_1, x_2, \mathbf{a}, s) dx_1 dx_2 = 0 \tag{2}$$

for any radial support Θ, such the wedge drawn in yellow in Fig. 4(c), including as maximal Θ the whole image domain Ω.

[1]The DoG is numerically negligible outside of the interval $(-18s, 18s)$, with the radius $18s$ being the *six-sigma* range for the standard deviation $3s$ of the wider Gaussian f_{3s}. All cases of practical relevance satisfy the inequality $d(c_1, c_2, \mathbf{a}) > 18s$ by a large margin, which means that tails of the DoG have no numerical relevance.

2.2. Integral cost function

Consequently, for an image z and an ellipse $E(\mathbf{a})$, the cost function is written as

$$\begin{aligned} C(z, \mathbf{a}, s) &= -\iint_{\Omega} z(x_1, x_2) g(x_1, x_2, \mathbf{a}, s) dx_1 dx_2 \\ &+ \lambda(\max(0, \frac{\max(r_1, r_2)}{\min(r_1, r_2)} - \mathrm{CI}))^2, \end{aligned} \tag{3}$$

where $\lambda = 0.5$ is a regularization parameter selected experimentally, and the term $\mathrm{CI} = 1.4$ is used to model the inverse of the minimal allowable cephalic index, see, e.g., [13] for a justification of this value. The regularization term is introduced to speed up the convergence of the minimization algorithm in those images where the skull is not fully visible (e.g. see the right-most panels in Fig. 2). Practically, the regularization term allows to reduce the number of the re-starts of the minimization algorithm by preventing the convergence of the algorithm to obviously incorrect solutions.

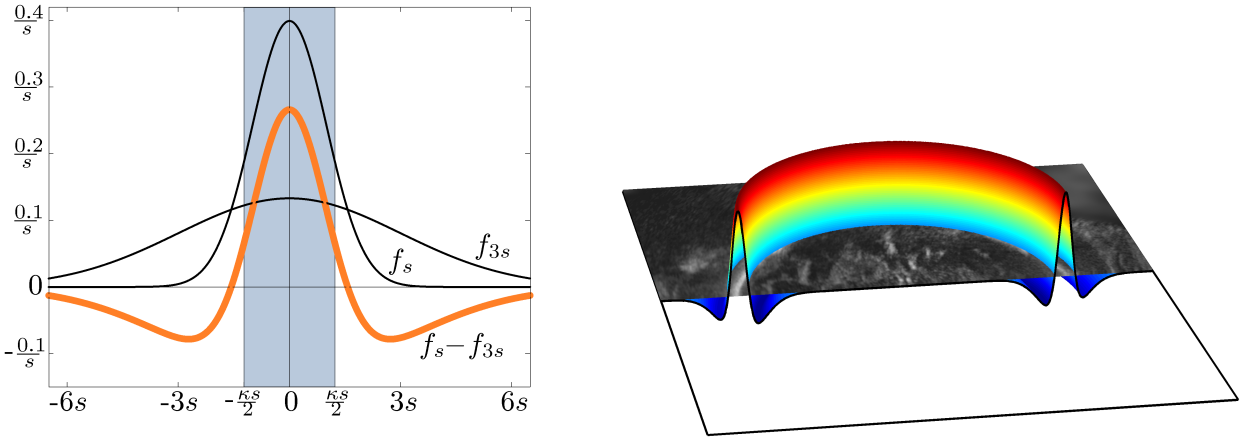
The cost function $C(z, \mathbf{a}, s)$ is affine with respect to z and, due to the characteristics of $g(x_1, x_2, \mathbf{a}, s)$, it is not affected by the presence of large uniform regions in the image. As observed in Eq. (2), the integral $\iint_{\Omega} g(x_1, x_2, \mathbf{a}, s) dx_1 dx_2 = 0$ for all reasonable $\mathbf{a}$ and s, therefore, if the image is noninformative (has a constant value in every pixel), every ellipse receives the score zero. This property is important because it ensures that an addition of a constant to the image intensities does not alter the value of the cost function, i.e. $C(z + B, \mathbf{a}, s) = C(z, \mathbf{a}, s)$, where B is a scalar and $z + B$ is to be interpreted as adding B to each intensity $z(x_1, z_2)$.

2.3. Discretization of Cost Function

For an $N_1 \times N_2$ image z, the integral in Eq. (3) is computed as the following discrete sum:

$$\begin{aligned} C(z, \mathbf{a}, s) &= -\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} z(i, j) g(i, j, \mathbf{a}, s) \\ &+ \lambda(\max(0, \frac{\max(r_1, r_2)}{\min(r_1, r_2)} - \mathrm{CI}))^2, \end{aligned} \tag{4}$$

where $z(i, j)$ denotes the intensity of z in the spatial position (i, j). Image padding is necessary for the zero-score condition (2) to hold for the noninformative images whenever the integral is approximated as in Eq. (4). We did not implement the image padding in the original challenge submission and had to compensate for it by introducing a multiplicative regularization term and an additional regularization parameter (γ in Eq. (1) of [9]). However, by implementing the image padding, we can safely eliminate this regularization and the corresponding parameter. Based on the same considerations given in Footnote 1, only a narrow innermost portion of the padding border is numerically significant with respect to the image model g. Our implementation uses a padding border 120-pixel wide, which was found to be sufficient in all experiments.



(a)

(b)

(c)

Figure 4: (a) Difference of Gaussians. The two univariate Gaussian probability density functions f_s and f_{3s} with standard deviation s and $3s$, respectively, and their difference $f_s - f_{3s}$. The central gray band depicts the interval $\left[-\frac{\kappa s}{2}, \frac{\kappa s}{2}\right]$ of width κs that corresponds to the skull thickness ($\kappa = 2.45$). The regions where the difference of Gaussians is negative capture instead the tissues having density lower than that of the skull, as illustrated in the panel (b), which shows a cross-section of the surface $g(x_1, x_2, \mathbf{a}, s)$ shown on the top of an ultrasound image. (c) The grid in the figure shows the coordinate system upon which the numerator of g is constructed. The grid is composed by radial lines stemming from the center of the ellipse, and by contour lines located at fixed radial distance from the ellipse (the distance is measured along the radial lines). The green patches in the figure illustrate two finite area elements; the denominator of g is proportional to their areas, making the integration of g over radial wedges (yellow) to be equivalent to integration of the DoG at the numerator over a rectilinear domain.

3. Methods

The DoGEll method works by minimizing the cost function described in Section 2 based on a modified multiscale, multistart Nelder-Mead algorithm (described in Section 3.2). However, we will start this section by describing the applied image preprocessing in Section 3.1. Finally, we will describe the computation of the biometric parameters based on the segmentation results in Section 3.3.

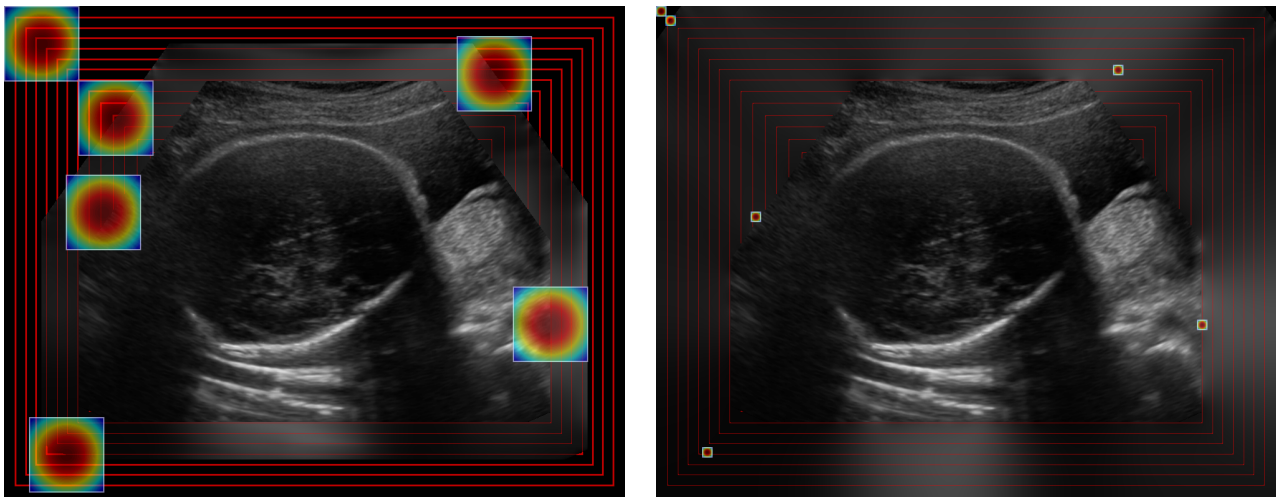
3.1. Preprocessing

To minimize the adverse effects caused by the heterogeneous contrast and the black background outside the field of view (see Fig. 2), we carry out the following two preprocessing operations on each unprocessed image, which we denote as z_{orig}. First, we extrapolate the image inside of the scanned area to fill the areas outside the field-of-view as well as a padding border which is appended to the four sides of the images. Without such an extrapolation, the boundary of the field of view might be confused with the boundary of skull that might cause the optimization algorithm in Section 3.2 to fail. This is done by a constrained iterative diffusion process illustrated in Fig. 5. Leaving the content within the field of view untouched, at each iteration we convolve the image with an isotropic smoothing kernel whose support has radius proportional to the maximum distance the between the already diffused image and the image boundary. In this way, as the diffused image reaches the boundary, the smoothing kernel approaches a Dirac delta. By doing so we achieve fast convergence of the diffusion process, as well as a smooth junction between the diffused image and the unmodified image in the scanned area. In particular, as illustrated in Fig. 5, we use a Kaiser 2-D kernel whose support radius is a quarter of the maximum distance between the already diffused image and the image boundary. About 18 iterations suffice for the diffusion to reach the boundary. We denote the extrapolated image as z_{extr}. Note that z_{extr} and z_{orig} coincide within the field of view.

The second preprocessing operation consists of a stabilization of the local contrast and intensities of z_{extr} leveraging discrete cosine transform (DCT) domain smoothing [21]. We achieve a smoothly varying local stabilization of the intensities and contrast of z_{extr} by 1) computing its DCT spectrum, 2) attenuating the low-frequency portion of the spectrum, and 3) by inverting the transform. In particular, the attenuation follows the usual zig-zag sorting of the coefficients [21]. This stabilized image is denoted as z (see Figure 6 for an example). The DCT-domain filtering is a practical tool (but not the only reasonable one) to perform a strong smoothing needed for the stabilization and a better alternative to the convolutive high pass filter that would require a large support and require more computational effort.

3.2. Optimization Algorithm

The cost function (3) is non-convex with respect to $\mathbf{a}$ and s and has several local minima. Therefore, we must use an algorithm which aims to find its global minimum to avoid the sensitivity to the algorithm initialization. We employ a multistart Nelder-Mead (NM) algorithm that repeats the following steps T times (T is defined below):



(a) (b)

Figure 5: The iterative diffusion process. At each diffusion iteration, we convolve the image with an isotropic smoothing kernel whose radius is proportional to the maximum distance between the image boundary and the previously diffused data. The image intensities within the field of view are preserved during the diffusion. As the diffusion progresses the data gets closer to the boundary and thus the smoothing kernel shrinks, yielding a seamless junction between the scanned and the extrapolated areas. The two subfigures show the intermediate results of diffusion (a) after the first and (b) after the eighth iteration, together with a few of the smoothing kernels. The red contour lines illustrate the Manhattan distance from the boundary.

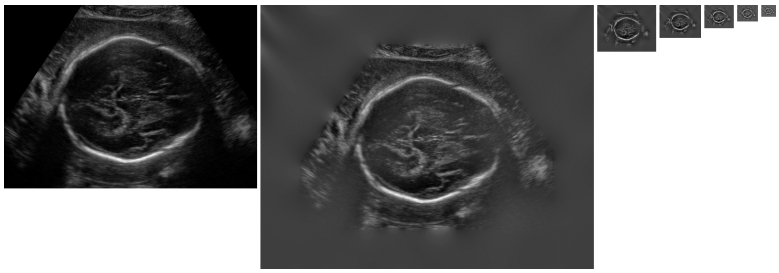


Figure 6: From left: z_{orig}, z, and lower scale versions of z $z_{(D)}$ for $D = [32/\sqrt{2}, 16, 16/\sqrt{2}, 8, 8/\sqrt{2}, 1]$. z is larger than z_{orig} because of the addition of the padding border.

1. generate an initial $\mathbf{a}$;
2. execute the NM minimization algorithm and if the minimum found is the best so far, save it;
3. apply a random perturbation to the best found solution, and go to step 2.

To further accelerate the convergence of the minimization algorithm, we follow a coarse-to-fine approach by first fitting the ellipse on a lower-resolution version of z and then using the found fit as the initial point on the higher-resolution image. We implement this approach recursively, using root-dyadic downscaling factors $D = [\sqrt{2}^9, \sqrt{2}^8, \sqrt{2}^7, \sqrt{2}^6, \sqrt{2}^5, \sqrt{2}^0]$, i.e., $D = [32/\sqrt{2}, 16, 16/\sqrt{2}, 8, 8/\sqrt{2}, 1]$ (see Fig. 6). The low-resolution image corresponding to the downscaling factor D is denoted as $z_{(D)}$. The multistart scheme is run for $T = 500, 75$, and 25 times for the three highest downscaling factors (the three lowest resolution images) $\sqrt{2}^9, \sqrt{2}^8, \sqrt{2}^7$, respectively, and only once for the other downscaling factors $(\sqrt{2}^6, \sqrt{2}^5, 1)$. These values are chosen empirically to produce a compromise between computational cost and reliability. For the initial ellipse, we use a small circle centered at the center of the image for the lowest resolution image and set the initial rotation angle as zero. For the subsequent resolution levels, the ellipse is initialized with the output of the previous resolution level. We emphasize that the algorithm initialization does not require any user intervention and the algorithm is not sensitive to its initialization.

Our implementation of the Nelder-Mead simplex search algorithm, originally proposed in [19], is based on [14]. However, we accelerate the implementation of [14] by incorporating an insertion sort algorithm. This modification is briefly explained next. The Nelder-Mead algorithm is a direct optimisation algorithm designed to iteratively find the minimiser of functions $f : \mathbb{R}^n \to \mathbb{R}$ (here, $n = 6$). At each iteration, the algorithm transforms the geometry of a $(n + 1)$-points simplex in the domain of f; in particular, the simplex can be reflected, expanded, contracted, or shrunk. Operatively, each iteration of the algorithm first evaluates f in each point of the simplex and optimizes the position of those having the larger values retaining the position of the best one. Then, the $n + 1$ points of the newly formed simplex are sorted based on their functional value, and the best n ones are used to compute a mean point. The shrinkage requires the computation of n new points, whereas the expansion, reflection, and contraction require the computation of just one new point. Thus, the latter cases, which are the most common in practice, can be accelerated by sorting the points of the simplex using an insertion sort algorithm because the old points of the simplex are already ordered [23]. Additionally, the new mean point can be computed by updating the old one by subtracting the removed point and adding the new one, instead of averaging all n points [23]. The coefficients of reflection, expansion, contraction, and shrinkage are as given in [14] for the standard Nelder-Mead algorithm.

3.3. Computation of biometric parameters

The major and minor axes of the ellipse directly correspond to the OFD and BPD, when measured from *center-to-center* of the skull, respectively. We denote these center-to-center measurements as

$$\text{OFD}_{\text{cc}} := 2\max(r_1, r_2), \qquad \text{BPD}_{\text{cc}} := 2\min(r_1, r_2).$$

To obtain *outer-to-outer* measurements OFD and BPD definitions, as in the manual analysis in the ISBI 2012 challenge data [22], we should increment the center-to-center measures by the calvarial thickness T:

$$\text{OFD} := \text{OFD}_{\text{cc}} + \text{T}, \qquad \text{BPD} := \text{BPD}_{\text{cc}} + \text{T}.$$

The thickness T is directly proportional to the standard-deviation parameter s of the fitted surface g. Therefore, after minimization of the cost functional, we can estimate the thickness as

$$\hat{\text{T}} = \kappa \hat{s},$$

where $\kappa = 2.45$ is the proportionality constant, as illustrated in Fig. 4 (a). This specific value of κ was determined through the correction formula of [7], which expresses the typical value of the skull thickness as a function of the BPD_{cc}:

$$\text{T}\left(\text{BPD}_{\text{cc}}\right) = 1.31 - 0.119\text{BPD}_{\text{cc}} + 0.00472\text{BPD}_{\text{cc}}^2 - 0.00003627\text{BPD}_{\text{cc}}^3. \tag{5}$$

where BPD_{cc} is expressed in millimeters. This formula provides an alternative estimate of the thickness, which depends on $2\min\left(\hat{r}_1, \hat{r}_2\right)$ and is independent of $\hat{s}$. The choice of κ must be such that the two alternative estimates of T are consistent. Therefore, given a dataset of ultrasound images, we define κ as

$$\kappa = \text{mean}_j \left\{ \frac{\text{T}\left(2\min\left(\hat{r}_1\left(j\right), \hat{r}_2\left(j\right)\right)\right)}{\hat{s}\left(j\right)} \right\},$$

where j is an index of the image in the dataset and the numerator gives the thickness estimate computed from the estimated minor axis through (5). We used the dataset described in section 4.1 to define κ. Note that this relies only on the automatic segmentations and we do not use any manually generated ground truth to define κ.

4. Experiments

4.1. Material

For the evaluation of the DoGEll method, we used the image data and manual analysis of the *Challenge US: Biometric Measurements from Fetal Ultrasound Images* of the IEEE International Symposium on Biomedical Imaging (ISBI) described in detail by [22]. However, we repeat the most important aspects of the used data here in order to make this paper self-contained. We use 90 ultrasound images of the fetal head acquired by trained clinicians using the same mid-range ultrasound machine Philips HD9 and following the protocols defined in [11]. The 90 images represented three different gestational ages (21, 28, and 33 weeks) with a total of 30 images per gestational age (90 different fetuses). For each gestational age, images were graded as being of low, medium, or high quality and were selected as objectively as possible to create real image data sets as used in clinical practice. The images were in an anonymized

DICOM format and automatically cropped to the size of 756×546 pixels. The pixel size varied between the images.

A total of three experts, with different degrees of expertise, participated in defining the fetal head ground truth. The ground truth consisted of 1) manually fitted ellipses to the fetal head and 2) standard clinical measurements of BPD, OFD and HC (derived from BPD and OFD) on each image. All experts fitted the ellipses and performed the measurements twice on each image. Therefore, altogether 540 (2 repetitions $\times$ 3 experts $\times$ 90 images) manual annotations were used in the experiments. The experts had the following levels of expertise:

- **Expert 1**: Clinician (fetal medicine specialist) with 10 year postgraduate experience in fetal US scans.
- **Expert 2**: Clinician (obstetrician) with 2 years experience in fetal US scans.
- **Expert 3**: Engineer with 1 year of experience.

4.2. Evaluation metrics

To report the quality of automatic segmentations in comparison to the manual ground truth, we have selected three measures from the original corpus of seven measures applied in the challenge [2].

We denote the sets of pixels inside the manually fitted ground truth ellipses by M and inside the automatically fitted ellipses by A. The Dice index [6] between the sets M and A is

$$\text{Dice}(A, M) = \frac{2|A \cap M|}{|A| + |M|}.$$

The Dice index measures the overlap of two different fetal head segmentations. It ranges between zero (no overlap) and one (perfect overlap), where higher Dice values indicate a better overlap. The Dice index evaluates the similarity of two segmented regions and is depictive of the overall segmentation quality. The Dice index is not especially sensitive to local differences between the segmentations and it is dimensionless. Therefore, it is complemented by two other evaluation metrics in this work.

Denote the contours of M and A by ∂M and ∂A, respectively. The maximum symmetric contour distance (MSD, also known as the Hausdorff distance) between the two contours is [26]

$$MSD(\partial A, \partial M) = \max(\max_{a \in \partial A} \min_{m \in \partial M} ||a - m||, \max_{m \in \partial M} \min_{a \in \partial A} ||a - m||).$$

[2] The quantitative results of all seven measures are provided in a supplement but we limit the discussion here to the 3 selected ones.

MSD equals 0 if the two contours are identical and a higher MSD indicates more pronounced differences between the contours. MSD is indicative for even very local differences between the two contours. The average symmetric contour distance (ASD) between the two contours is

$$ASD(\partial A, \partial M) = \frac{\int_{a\in\partial A} \min_{m\in\partial M} ||a-m|| + \int_{m\in\partial M} \min_{a\in\partial A} ||a-m||)}{|\partial A| + |\partial M|}.$$

ASD equals 0 if the two contours are identical and a higher ASD indicates more pronounced differences between the contours.

We further subjected the values of these metrics to statistical testing to find if there were significant differences between intra- and inter-expert segmentation variabilities and the automatic segmentation accuracy. We used the standard two-tailed t-test for hypothesis testing.

We compared the biometric parameters of interest derived based on the automatic segmentation to manual measurements. The criteria were the average difference $\bar{d}$ (that can be considered as a measure of bias) and its standard deviation s_d between the automatic and manual biometric measures; the manual measure for each fetus was selected to be the average of the two measurements of an expert to increase its accuracy against intra-expert variability. These quantities are referred to as the Bland-Altman plots by [22] (see [1]). Based on these two quantities, we can compute the root mean squared error $RMSE = \sqrt{\bar{d}^2 + s_d^2}$. $RMSE$ provides a single indicator of the difference in the derived biometrics that simplifies the method comparisons. Note that the point made in [1] against the use of correlation as a measure of agreement is particularly relevant here, where the data is such that the biometrics to be determined vary considerably due to different gestational ages of the fetuses. Therefore, we do not provide correlation coefficients between automatic and manual biometrics.

5. Results

5.1. Computation time

The computation time of a Matlab-based implementation of the algorithm was on average 4.62 seconds per image on a laptop (Intel(R) Core(TM) i5-3320M dual core CPU running at 2.60 GHz with 64-bit Windows 7 operating system, Matlab version R2013a). The maximal computation time per image was 5.43 seconds. These timings include the time spent for preprocessing of Section 3.1, cost function optimization of Section 3.2, and computation of biometric parameters of Section 3.3. A significant portion of the computation time (on average 3.62 seconds) was spent on the cost function minimization. Preprocessing required approximately 1 second per image.

The current Matlab-based implementation is sequential and single-threaded, where the Nelder-Mead algorithm is implemented in C and compiled into a mex file, and the other parts of the method are written in Matlab. Further improvements in computation time would be gained by exploiting a multicore implementation and by a more significant usage of the more efficient C language. In particular, the

early stages of the optimization involving the multistart strategy can be parallelized in a straightforward manner.

5.2. Quantitative validation

Table 1: Quantitative evaluation measures. The top three rows provide the segmentation accuracy measures of two versions of DoGEll and the method of [24] (ranked second in the segmentation challenge after DoGEll) averaged over all 540 expert segmentations (6 segmentations of each of the 90 images). The values not involving DoGEll method are reproduced from [22].

Method	**Dice** (%)	**MSD** (mm)	**ASD** (mm)
DoGEll	97.73 ± 0.89	2.26 ± 1.47	0.91 ± 0.47
DoGEll (slow, [9])	97.80 ± 1.04	2.16 ± 1.44	0.88 ± 0.53
[24]	97.23 ± 0.77	2.59 ± 1.14	1.07 ± 0.39

Table 2: Comparison of DoGEll and manual analysis. For the reference, the intra- and inter-expert accuracy measures are given, see [22] for details. The rows "DoGEll vs Expert" provide performance metrics when the automatic method was compared to the manually fitted ellipses by a single expert. The values not involving DoGEll method are reproduced from [22].

Method	**Dice** (%)	**MSD** (mm)	**ASD** (mm)
DoGEll vs Expert 1	97.89 ± 0.95	2.12 ± 1.55	0.84 ± 0.47
DoGEll vs Expert 2	97.54 ± 0.99	2.38 ± 1.61	0.99 ± 0.51
DoGEll vs Expert 3	97.77 ± 1.08	2.29 ± 1.56	0.90 ± 0.55
Inter-expert 1 vs 2	97.87 ± 0.73	2.11 ± 1.12	0.86 ± 0.39
Inter-expert 2 vs 3	97.66 ± 0.77	2.24 ± 1.19	0.93 ± 0.42
Inter-expert 1 vs 3	97.83 ± 0.78	2.09 ± 0.99	0.86 ± 0.40
Intra-expert 1	98.24 ± 0.71	1.72 ± 0.81	0.69 ± 0.32
Intra-expert 2	98.28 ± 0.76	1.74 ± 1.09	0.68 ± 0.35
Intra-expert 3	98.01 ± 0.94	1.85 ± 1.10	0.79 ± 0.44

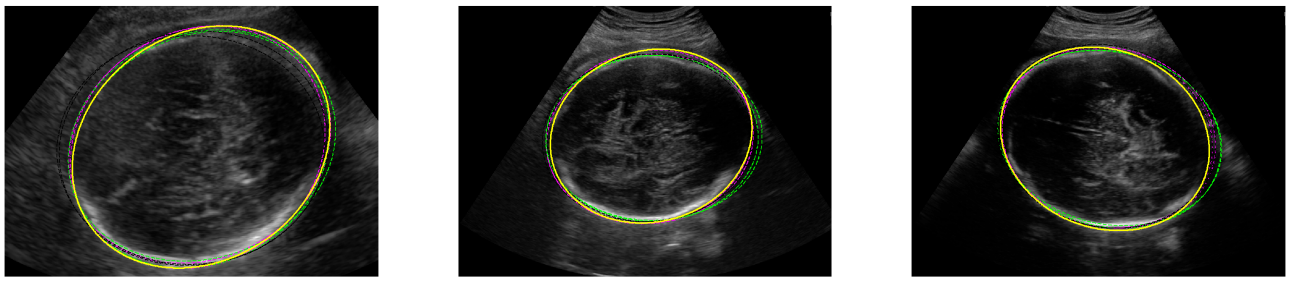
The quantitative evaluation results of the DoGEll are provided in Tables 1 and 2. Table 1 also provides the corresponding results of the earlier (slower) implementation of DoGEll, which won the ultrasound segmentation challenge [22], and the second best method [24] in the same segmentation challenge. Table 2 presents more detailed results of comparison between DoGEll and manual delineations. To simplify

the comparison to the manual segmentation accuracy, we have also reproduced the intra and inter-expert measures from [22] in Table 2. Note that all the results corresponding to [9] are given only for comparison purposes and the discussion about the DoGEll refers to the fast implementation as presented in this work. The results can be summarized as follows.

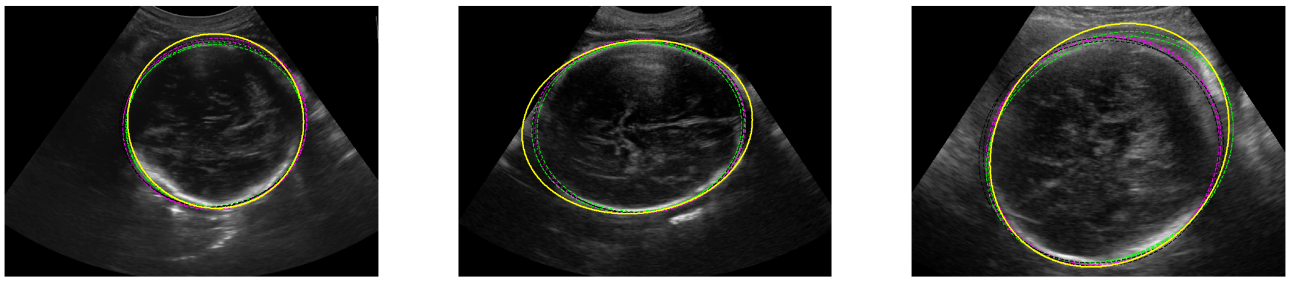
1. The segmentation accuracy of DoGEll was significantly better than that of any other segmentation method participating to *Challenge US: Biometric Measurements from Fetal Ultrasound Images*. In particular, the difference between the segmentation accuracy of DoGEll and the second best approach in the challenge [24] was significant (one-sided $p < 0.05$ for all three measures (Dice, MSD, and ASD)). This was computed only using information from [22] which necessitated the use of an unpaired t-test. The differences in average performance between the slow version of DoGEll in [9] (computation time of 5 min per image) and the fast implementation in this paper (computation time of under 5 seconds per image) were not significant at the alpha-level $p = 0.05$ ($p > 0.3$ for all three measures).
2. The average segmentation accuracy of the automatic DoGEll method was similar to the inter-expert variability of the manual segmentations with all three measures. More specifically, the average Dice coefficient and ASD of DoGEll were better than the corresponding average inter-expert measures between experts 2 and 3. For other cases, the average measures of DoGEll were slightly worse than the inter-expert variability (but not significantly, one-sided $p > 0.05$ for all expert pairs and validation metrics).
3. The segmentation accuracy of DoGEll was worse than the intra-rater variability (one-sided $p < 0.05$ in all cases) as expected. However, it is worth to note that the inter-expert variabilities were also greater than intra-expert variabilities.

An interesting aspect was revealed when comparing the measures between our method and the segmentation by Expert 1 (DoGEll vs Expert 1) - who is the most experienced one - versus the inter-rater variabilities between Expert 1 and other experts (the rows Inter-expert 1 vs 2 and Inter-expert 1 vs 3 in Table 2). If we assume that the ground truth segmentation was the one by Expert 1 then our automatic method would, on average, outperform the manual delineations of the Experts 2 and 3 in terms of Dice index or ASD. However, it must be noted that differences were very small (insignificant at the $p < 0.05$ level) and, if using MSD to judge performance, the Experts 2 and 3 would slightly outperform our method.

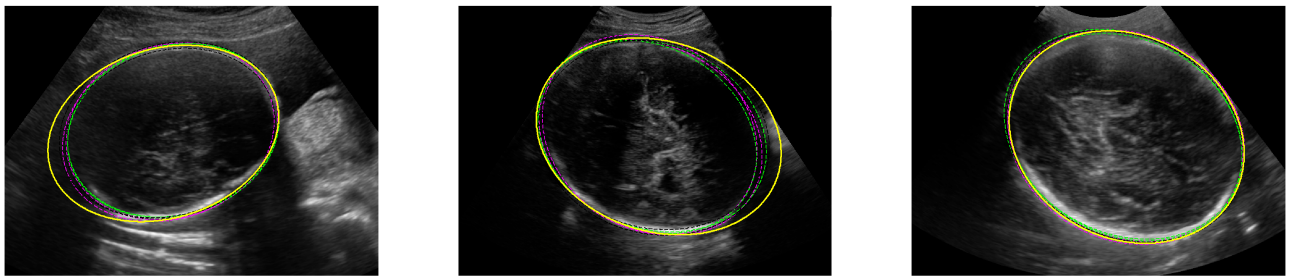
It is interesting to study the worst cases in terms of segmentation accuracy. These are best identified by the MSD metric, which is sensitive to very local differences between the two segmentations (automatic and manual). Eight worst cases of the total 90 in terms of the MSD averaged over all expert segmentations are shown in Fig. 7. Figure 7 shows that the head contour by the automatic segmentation mostly was in the space set by expert segmentations, and only very few clear local differences from all expert segmentations can be observed. Figure 7 also shows the image with the median (46th) MSD for reference in the bottom right panel.



MSD 3.97 mm [2.10 mm, 7.76 mm] 4.06 mm [2.38 mm, 6.19 mm] 4.11 mm [2.08 mm, 6.46 mm]



MSD 4.70 mm [3.54mm, 6.17 mm] 6.11 mm [5.32 mm,7.44 mm] 6.98 mm [3.40 mm, 9.10 mm]



MSD 7.95 mm [5.88 mm, 9.46 mm] 9.18 mm [6.24mm, 12.10 mm] 1.79 mm [1.01mm, 3.25mm]

Figure 7: Eight worst cases and the average case (the bottom right panel) of the total 90 in terms of the MSD averaged over all the expert segmentations. The outer head contour by the automatic method is shown in yellow and outer head contours by the experts are shown in magenta (Expert 1), green (Expert 2), and black (Expert 3). Below each image, the average MSD over all expert segmentations is reported followed by the range of MSDs, i.e., the minimum and maximum MSD between the automatic segmentation and six different expert segmentations.

The mean values (biases) and standard deviations of the differences between the manual and automatic biometric measurements are reported in Table 3. Table 3 also provides inter-observer variabilities reproduced from [22] and the corresponding RMSEs computed based on these. A few observations can be made based on these values.

1. The average differences ($\bar{d}$) for the automatic measures by DoGEll were larger than between expert average differences. Moreover, $\bar{d}$ for DoGEll was always negative meaning that, on average, biometric parameters as estimated by DoGEll were slightly smaller than the corresponding parameters estimated by experts. However, if this bias is considered to be important, it can be easily corrected by tuning the parameter κ of Section 3.3. We here tuned the parameter κ based on automatic segmentations, but to reduce the bias, it could be tuned by manual segmentations. However, in this case, the method development and evaluation would not be independent and we wanted to avoid this.
2. The standard deviations of the differences (s_d) can be argued to be more important than the average differences ($\bar{d}$) because they cannot be improved by simple parameter tuning (unlike the average differences). In Table 3, s_d for the automatic measurements as compared to between expert s_d were (on average) smaller for the BPD, larger for the OFD, and almost equal for the HC. For example, treating again the measurements of the most experienced expert (E1) as a gold standard and s_d as a performance index, the automatic method outperformed the Expert 2 in the measurement of BPD ($s_d = 1.25mm$ vs. $s_d = 1.66mm$, F-test between the two s_d^2 values p < 0.01, one-sided), the expert 2 outperformed the automatic method in the measurement of OFD ($s_d = 3.12mm$ vs. $s_d = 2.36mm$, F-test p <0.01, one-sided) and, for HC, the difference in s_d^2 was non-significant ($s_d = 4.71mm$ vs. $s_d = 4.16mm$), according to F-test at the alpha level $p = 0.05$).
3. RMSEs for the automatic method were lower for the BPD measures than the inter-observer RMSEs between Expert 1 and Experts 2 or 3. Instead, the BPD measurements of Experts 2 and 3 were well in line, consistently resulting in a RMSE of only 1.00 mm, which is less than the corresponding RMSEs for DoGEll (1.37 mm and 1.54 mm). For OFD and HC, the RMSEs by the automatic method were slightly higher than inter-expert RMSEs.

6. Discussion

We have presented an automatic method, DoGEll, for the segmentation of the fetal skull from 2-D ultrasound images. The DoGEll method is based on a global optimization of a novel cost function by a multi-start multi-scale Nelder-Mead algorithm. The cost function is based on the assumption that the cross-section of the skull has a roughly elliptical shape. The main difference between our cost function and previous ellipse-fitting methods [10, 17, 27] are that in our cost function the skull has a finite thickness instead of being modeled with an ellipse contour as in and our cost function models the image intensity also around the skull. Moreover, we do not consider the calvarial thickness to be a user defined constant

Table 3: The accuracy of BPD, OFD and HC measurements. All values are expressed in millimeters. Expert 1 is abbreviated as E1. The columns ' vs. all experts' provide the average performance of automatic methods with respect to all expert measurements. The column '[24] vs. all experts' refers to the average performance of the second best method [24] after DoGEll in the ISBI segmentation challenge [22]. For BPD, the method of Sun [25] achieved quantitative values closer to the manual ground truth ($\bar{d} = 0.58mm, s_d = 1.24mm, RMSE = 1.37mm$) than DoGEll. The RMSEs for [24, 25] were computed based on $\bar{d}, s_d$ values presented in [22].

Measurement	DoGEll vs. E1			DoGEll vs. E2			DoGEll vs. E3			DoGEll vs. all experts		
	$\bar{d}$	s_d	$RMSE$	$\bar{d}$	s_d	$RMSE$	$\bar{d}$	s_d	$RMSE$	$\bar{d}$	s_d	$RMSE$
BPD	-1.01	1.25	1.60	-1.22	0.95	1.54	-0.84	1.09	1.37	-1.02	0.97	1.41
OFD	-1.72	3.12	3.56	-0.27	3.16	3.17	-0.62	2.98	3.04	-0.87	2.84	2.97
HC	-2.25	4.19	4.76	-2.99	4.43	5.34	-1.76	3.79	4.19	-2.34	3.72	4.39
	E1 vs. E2			E2 vs. E3			E1 vs. E3			[24] vs. all experts		
BPD	0.39	1.66	1.71	-0.47	0.89	1.01	-0.08	1.84	1.84	-1.65	0.93	1.89
OFD	-1.55	2.36	2.82	1.09	2.75	2.96	-0.45	2.24	2.28	-0.96	2.92	3.07
HC	0.68	4.15	4.20	0.65	3.76	3.83	1.33	4.07	4.28	-3.46	4.06	5.33

as in [12] but as a hyper-parameter to be optimized along with the five ellipse parameters. This means that we do not fit an ellipse or some other contour to an image as in [10, 17, 27, 12] but, instead, we first construct a template image of the fetal skull based on the ellipse parameters and then we match the resulting model with the observed ultrasound image. In this way, we are able to circumvent thresholding and skeletonization operations usually required by Hough transform based methods [10, 17, 27]. Also, our image model is different from the simple two-component (skull and background) mixture model of the region based segmentation method in [12] and we avoid the manual initialization in [12]. As a result, the DoGEll method is fully automatic and, at the same time, extremely robust against image imperfections as demonstrated by experiments presented in this paper.

We have presented a quantitative validation of the DoGEll method based on 90 images from a recent ultrasound segmentation challenge [22]. In particular, the automatic segmentations and the biometric parameters (BFD, OFD, and HC), derived from the segmentations, were validated against the corresponding manual analysis provided by three experts. The images for the challenge were purposefully selected so that the image quality varied between the images. We have shown that, even for the low quality images, our method yielded segmentations with an accuracy comparable to inter-expert variability of segmentations. We statistically tested if the accuracy differences between any two analysis methods were significant. However, we should point out that a statistically significant difference between the measurements by two different analysis 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.

Only few works on automated segmentation of fetal ultrasound images have addressed the accuracy of head segmentations against manual segmentations. Carneiro et al. [2] reported mean Hausdorff distances (MSDs) of 4.83 mm (their set 1) and 4.15 mm (their set 2) and average distances (ASDs) of 3.39 mm and 2.76 mm when comparing automatic head contour delineations to the delineations by experts. Chalana and Kim [3] reported mean Hausdorff distance of 4.64 mm and the average distance of 2.09 mm. The corresponding error measures for our method, listed in Table 1, were: an average MSD of 2.26 mm and the average ASD of 0.91 mm. Thus, in the light of this comparison, the performance of our method appears superior to these two albeit it must be stressed that the validation results may not be comparable because of the different sets of images used. A majority of the reports on automated fetal ultrasound segmentation have provided quantitative validation in the terms of the derived biometric parameters. For example, Yu et al. [27] reported mean absolute difference (MAE) of 2.5 mm for BPD and 5.7 mm for HC between automatically and manually derived biometric parameters. Noting that the RMSE is always greater than MAE, the biometric parameters by our method appeared more accurate than those reported in [27]; however, we must repeat the earlier note that the validation results cannot be directly compared. In particular, the gestational ages of the fetuses studied in [27] were somewhat higher than those in this paper.

Although comparisons to the previous methods in literature are problematic due to different datasets used for validation and a poor availability of the software implementations of the methods, the segmentation accuracy by our method can be directly compared to the results of other methods which participated in the Challenge US: Biometric Measurements from Fetal Ultrasound Images held in conjunction of the ISBI 2012 conference [22] [3]. As already noted, the segmentation accuracy of our method was superior to that of the other methods which participated in the Challenge (see Section 5.2). Also, when using RMSE as the performance measure, the biometric measurements by our method were the most accurate for OFD and HC (all experts RMSEs 2.97 mm (OFD) and 4.39 mm (HC) against RMSEs of 3.07 mm and 5.33 mm of [24] which was the second best method in the challenge in this respect), and the second most accurate for BPD (RMSE of 1.41 mm compared to RMSE of 1.37 mm of [25]). However, having a small standard deviation s_d can be considered more important than having a small bias $\bar{d}$ and s_d of our method was smaller than that of [25] for all three biometric measurements. The RMSEs for [24, 25] were computed based on $\bar{d}, s_d$ values presented in [22].

Finally, the source-code of DoGEll is available at `http://www.cs.tut.fi/~foi/dogell/` under an license permitting free non-commercial use of it.

[3] The Challenge data will be made available to the research community. The estimated date for data release to the research community is anticipated to be in autumn 2014. The challenge website `http://www.ibme.ox.ac.uk/challengeus2012` will then be updated to allow new segmentation results to be uploaded for evaluation and comparison to previous methods.

Acknowledgments

The work of A. Foi, M. Maggioni, A. Pepe and J. Tohka was supported by the Academy of Finland under grants 130275 and 252547, and by Tampere Graduate School in Information Science and Engineering (TISE). S. Rueda and J.A. Noble were supported by the Wellcome/EPSRC Centre of Excellence in Medical Engineering - Personalized Healthcare, WT 088877/7/09/Z. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

[1] Bland, J., Altman, D., 1986. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 327, 307–310.

[2] Carneiro, G., Georgescu, B., Good, S., Comaniciu, D., 2008. Detection and measurement of fetal anatomies from ultrasound images using a constrained probabilistic boosting tree. IEEE Trans Med Imaging 27, 1342–1355. URL: http://dx.doi.org/10.1109/TMI.2008.928917, doi:10.1109/TMI.2008.928917.

[3] Chalana, V., Kim, Y., 1997. A methodology for evaluation of boundary detection algorithms on medical images. IEEE Trans Med Imaging 16, 642–652. URL: http://dx.doi.org/10.1109/42.640755, doi:10.1109/42.640755.

[4] Chalana, V., Winter, T., Cyr, D.R., Haynor, D.R., Kim, Y., 1996. Automatic fetal head measurements from sonographic images. Acad Radiol 3, 628–635.

[5] Chervenak, F.A., Skupski, D.W., Romero, R., Myers, M.K., Smith-Levitin, M., Rosenwaks, Z., Thaler, H.T., 1998. How accurate is fetal biometry in the assessment of fetal age? Am J Obstet Gynecol 178, 678–687.

[6] Dice, L., 1945. Measures of the amount of ecologic association between species. Ecology 26, 297 – 302.

[7] Doubilet, P., Greenes, R., 1984. Improved prediction of gestational age from fetal head measurements. Am J Roentgenol 142, 797 – 800.

[8] Dudley, N.J., 2005. A systematic review of the ultrasound estimation of fetal weight. Ultrasound Obstet Gynecol 25, 80–89. URL: http://dx.doi.org/10.1002/uog.1751, doi:10.1002/uog.1751.

[9] Foi, A., Maggioni, M., Pepe, A., Tohka, J., 2012. Head contour extraction from fetal ultrasound images by difference of gaussians revolved along elliptical paths, in: Proceedings of Challenge US: Biometric Measurements from Fetal Ultrasound Images.

[10] Hanna, C.W., Youssef, A.B.M., 1997. Automated measurements in obstetric ultrasound images, in: ICIP97, pp. 504 – 507.

[11] International Fetal and Newborn Growth Consortium, 2008. The international fetal and newborn growth standards for the 21st century (intergrowth-21st) study protocol. www.intergrowth21.org.uk.

[12] Jardim, S., Figueiredo, M.A.T., 2005. Segmentation of fetal ultrasound images. Ultrasound Med Biol 31, 243–250. URL: http://dx.doi.org/10.1016/j.ultrasmedbio.2004.11.003, doi:10.1016/j.ultrasmedbio.2004.11.003.

[13] Kurmanavicius, J., Wright, E., Royston, P., Wisser, J., Huch, R., Huch, A., Zimmermann, R., 1999. Fetal ultrasound biometry: 1. head reference values. Br J Obstet Gynaecol 106, 126 – 135.

[14] Lagarias, J., Reeds, J.A., Wright, M.H., Wright, P.E., 1998. Convergence properties of the nelder-mead simplex method in low dimensions. SIAM Journal of Optimization 9, 112 – 147.

[15] Loughna, P., Chitty, L., Evans, T., Chudleigh, T., 2009. Fetal size and dating: charts recommended for clinical obstetric practice. Ultrasound 17, 161 – 167.

[16] Lourens, T., 1995. Modeling retinal high and low contrast sensitivity filters, in: From Natural to Artificial Neural Computation. Springer, pp. 61–68.

[17] Lu, W., Tan, J., Floyd, R., 2005. Automated fetal head detection and measurement in ultrasound images by iterative randomized hough transform. Ultrasound Med Biol 31, 929–936. URL: http://dx.doi.org/10.1016/j.ultrasmedbio.2005.04.002, doi:10.1016/j.ultrasmedbio.2005.04.002.

[18] Marr, D., Hildreth, E., 1980. Theory of edge detection. Proceedings of the Royal Society of London. Series B. Biological Sciences 207, 187–217.

[19] Nelder, J., Mead, R., 1965. A simplex method for function minimization. Computer journal 7, 308 – 313.

[20] Pathak, S.D., Chalana, V., Kim, Y., 1997. Interactive automatic fetal head measurements from ultrasound images using multimedia computer technology. Ultrasound Med Biol 23, 665–673.

[21] Rao, K., Yip, P., 1990. Discrete cosine transform: algorithms, advantages and applications. Academic Press.

[22] Rueda, S., Fathima, S., Knight, C., Yaqub, M., Papageorghiou, A.T., Rahmatullah, B., Foi, A., Maggioni, M., Pepe, A., Tohka, J., Stebbing, R., McManigle, J., Ciurte, A., Bresson, X., Cuadra, M.B., Sun, C., Ponomarev, G.V., Gelfand, M.S., Kazanov, M.D., Wang, C.W., Chen, H.C., Peng,

C.W., Hung, C.M., Noble., J., 2014. Evaluation and comparison of current fetal ultrasound image segmentation methods for biometric measurements: A grand challenge. IEEE Trans Med Imaging 33, 797 – 813.

[23] Singer, S., Singer, S., 2004. Efficient implementation of the nelder-mead search algorithm. Appl. Numer. Anal. Comput. Math. 1, 524 – 534.

[24] Stebbing, R.V., McManigle, J.E., 2012. A boundary fragment model for head segmentation in fetal ultrasound., in: Proceedings of Challenge US: Biometric Measurements from Fetal Ultrasound Images, ISBI 2012, pp. 9 – 11.

[25] Sun, C., 2012. Automatic fetal head measurements from ultrasound images using circular shortest paths, in: Proceedings of Challenge US: Biometric Measurements from Fetal Ultrasound Images.

[26] Vergeest, J., Spanjaard, S., Song, Y., 2003. Directed mean hausdorff distance of parameterized freeform shapes in 3d: a case study. Vis. Comput. 19, 480 – 492.

[27] Yu, J., Wang, Y., Chen, P., 2008. Fetal ultrasound image segmentation system and its use in fetal weight estimation. Med Biol Eng Comput 46, 1227–1237. URL: http://dx.doi.org/10.1007/s11517-008-0407-y, doi:10.1007/s11517-008-0407-y.