



Segmentation of embryonic and fetal 3D ultrasound images based on pixel intensity distributions and shape priors



Sonia Dahdouh^{a,*}, Elsa D. Angelini^a, Gilles Grangé^b, Isabelle Bloch^a

^a Institut Mines-Telecom, Telecom ParisTech, CNRS LTCI, 46 rue Barrault, 75013 Paris, France

^b Maternité Port Royal AP-HP, Groupe Hospitalier Cochin Saint Vincent de Paul, 53 Avenue de l'Observatoire, 75014 Paris, France

ARTICLE INFO

Article history:

Received 10 June 2014

Revised 16 December 2014

Accepted 18 December 2014

Available online 10 January 2015

Keywords:

3D ultrasound

Obstetric imaging

Multi-phase level-set segmentation

Anatomical models

Statistical shape priors

ABSTRACT

This paper presents a novel variational segmentation framework combining shape priors and parametric intensity distribution modeling for extracting the fetal envelope on 3D obstetric ultrasound images. To overcome issues related to poor image quality and missing boundaries, we inject three types of information in the segmentation process: tissue-specific parametric modeling of pixel intensities, a shape prior for the fetal envelope and a shape model of the fetus' back. The shape prior is encoded with Legendre moments and used to constraint the evolution of a level-set function. The back model is used to post-process the segmented fetal envelope. Results are presented on 3D ultrasound data and compared to a set of manual segmentations. The robustness of the algorithm is studied, and both visual and quantitative comparisons show satisfactory results obtained by the proposed method on the tested dataset.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Since its introduction in the early 1980s, ultrasound (US) is used to monitor fetal growth during pregnancy (Bottomley and Bourne, 2009). In clinical routine, it is used to determine fetal age, assess fetal development and diagnose possible pathologies such as fetal aneuploidy, Down's syndrome or fetal growth restriction (Sonek, 2007). These assessments are based on a set of biometric measurements such as the head circumference, the nuchal translucency or the crown-rump length (CRL). As an illustration, the CRL (Robinson and Fleming, 1975) is used during the first trimester for pregnancy dating, and has been shown to be more reliable than the count of the number of weeks of amenorrhea (WA) (Campbell et al., 1985). Precise pregnancy dating is critical in multiple contexts such as: Down's syndrome screening, labor induction decision for post-term pregnancies and handling of preterm deliveries (Roberts and Thilaganathan, 2007). Regarding the need for precise biometric measures, small embryonic size or discordant growth of embryonic structures is associated with abnormal fetal development and poor vital prognosis (Bottomley and Bourne, 2009). Several studies have also investigated the relationship between small embryonic CRL and miscarriage (Reljic, 2001) and between altered embryonic growth and fetal growth restriction in the second and third trimesters of pregnancy

(Bottomley and Bourne, 2009). It seems that the single CRL observation is not enough to accurately predict miscarriage occurrence and efforts are made to propose new statistical models to improve miscarriage predictions (Abdallah et al., 2011; Bottomley et al., 2009).

In the past 10 years, 3D ultrasound has become an essential tool for fetal structures visualization (Levaillant, 2006), and embryonic and fetal volume (FV) measurements. Three-dimensional US-based biometric measures have been proposed in several studies to evaluate miscarriage probability and low birth weight (Aviram et al., 2004; Falcon et al., 2005). However, despite being related to birth weight, contradictory results regarding the relationship between fetal volume and preterm birth or growth impairment have been published (Smeets et al., 2013; Martins et al., 2008, 2009).

In many of these clinical studies, authors used the semi-automatic software tool VOCAL (General Electric Healthcare), to segment 3D anatomical structures (Smeets et al., 2013; Martins et al., 2009), which requires cumbersome manual supervision. Also, since this software cannot segment complex objects, limbs are usually excluded from the segmentation of the fetal envelope, when performed on fetuses older than 10 weeks of amenorrhea (Martins et al., 2009), while they are included in the earlier stages of pregnancy. This can lead to some inconsistency in longitudinal analysis of growth, since limbs represent a significant portion of the embryo or fetal volume (Blaas et al., 2006).

Regarding previous work on fetal ultrasound segmentation, only few of them dealt specifically with the problem of the segmentation of the fetus envelope on 3D US data, while many

* Corresponding author.

E-mail address: dahdouh@telecom-paristech.fr (S. Dahdouh).

concentrated on the automatic determination of biometry markers (Yu et al., 2008; Carneiro et al., 2008) or specific body parts such as the head or femur. In Jardim and Figueiredo (2005), for example, authors used a parametric deformable model and Rayleigh intensity distribution to segment fetal heads and femurs. A method combining region growing and an augmented reality system was proposed by Rousian et al. (2010), to alleviate the manual contouring task. However, manual initialization of seed points and pre-processing was required to manually erase the umbilical cord and the vitelline duct insertion in the yolk sac, and separate the fetus from the maternal tissues.

In order to overcome this issue, many authors have proposed to model the envelope of the RF-envelope signal using various statistical distributions such as Rayleigh (Noble and Boukerroui, 2006) or the k -distribution (Shankar, 1995). In Eltoft (2006), a modeling using a Rician inverse Gaussian distribution was made and compared to the k -distribution and a generalized Nakagami distribution.

One of the main difficulties when detecting structures in US images is their susceptibility to speckle. For US-based tissue segmentation, the envelope of the radio-frequency (RF) signal was modeled using a Nakagami distribution in Namburete et al. (2013), Destrempe et al. (2009), and Tsui et al. (2010), having the main advantage of being able to emulate other distributions by varying its shape parameter. It was successfully integrated into an Adaboost learning framework to detect anatomical landmarks in 2D fetal sonograms in Namburete et al. (2013). Few work has been dedicated to the specific statistical characterization of fetal and maternal tissues. In Anquez et al. (2013), the amniotic fluid was modeled using an exponential or a Rayleigh distribution, depending on whether the image was saturated or not. In Paragios et al. (2005), Normal distributions were used to segment echocardiographic images.

Deformable models and level-set implementations are a popular choice for the partition of US images into statistically homogeneous tissues. In Sarti et al. (2005), a maximum likelihood formalism was introduced to optimize, within a level-set framework, intensity homogeneities, modeled using Rayleigh distributions. In Anquez et al. (2013), a similar level-set formalism was exploited on 3D ultrasound data, but using tissue-specific intensity distributions to separate the amniotic fluid (fitted with an exponential or Gaussian distribution), from maternal and fetal tissues (fitted with a Rayleigh distribution). The authors also introduced the Gamma distribution as a generic model that can be used for all tissue types.

In this work, we refine the segmentation framework introduced in Anquez et al. (2013) by integrating a shape constraint into the 3D variational segmentation formulation, and by extending the framework to segment more than two tissue types. This allows us to achieve a first separation between maternal and fetal tissues. The shape constraint exploits a database of learned shape models, encoded with Legendre moments. Such a moment-based shape encoding was used in Foulonneau et al. (2009), Wojak et al. (2010), and Zhang et al. (2011) for different anatomical structures.

While the segmentation framework was able to separate the amniotic fluid from the utero-fetal tissues, separation of the fetus from the uterus wall was not always completely achieved. Therefore, a new post-processing step exploiting a set of back models is introduced to separate the fetus from connected structures such as the uterus wall. In this paper, we extend our previous work (Dahdouh et al., 2013) by achieving an automatic determination of the tissues distribution and refining the way shape priors are taken into account in the evolution of the deformable model. The quality of the segmentation of the fetal envelope was evaluated quantitatively on fifteen 3D US volumes, comparing to manual tracings and including inter-expert segmentation variability. A parameters sensitivity study was also conducted. Finally, the approach was extended for

the segmentation of internal fetal structures, with preliminary results presented.

2. Images and method

2.1. Image database

To design our method and our shape models, we exploited a database of eighteen 3D US volumes acquired during the first trimester of pregnancy, and including embryos (8, 9, 10 WA) and fetuses (12, 13 WA). For the rest of this paper, we use the term fetus (and fetal) to designate both embryos and fetuses. The data sets were provided by the Beaujon AP-HP hospital (France), and were acquired using a Voluson 730 Expert system (GE Healthcare) with a 3.7–9.3 MHz trans-vaginal volumetric probe. RF lines were acquired without compounding but using harmonic imaging. They were then post-processed and aggregated by the ultrasound scanning system to generate 3D US volumes with an isotropic voxel size¹ ranging from 0.21 to 0.96 mm³. Post-processing involves intensity normalization, and saturation of the lowest range of values.

An additional volume of a 22 WA fetus was provided by Philips Healthcare Research Labs (France). It was acquired using an iU22 transducer (Philips Ultrasound) with a 2–6 MHz volumetric probe. The image presents anisotropic voxels with a voxel size of 0.95 × 0.6 × 1.37 mm³. A more detailed description of this database can be found in Anquez et al. (2013).

2.2. Construction of the shape models database

The proposed segmentation method exploits a shape prior for the extraction of the fetal envelope from 3D US volumes. Unlike what has been done in Dahdouh et al. (2013) where the shape database covered only few ages and limbs positions, the design of our shape prior is based on the construction of an extended database of shapes encoding variability with respect to fetal age and limbs positioning.

First, the following prototypes of shape were gathered for the following gestational ages:

1. **7–10 WA:** Eight 3D prototype shapes were provided by the Multi-Dimensional Human Embryo project.²
2. **11–12 WA:** None.
3. **13 WA:** The 14 WA prototype was extrapolated, using the method described in Serrurier et al. (2013) and reference curves of biometric measures reported in Serrurier et al. (2013).
4. **14 WA:** We used the skeleton-based shape prototype described in Serrurier et al. (2013).

Since no prototype shapes were available at gestational ages 11–12 WA, we will assume that shapes from nearby ages will be sufficiently similar to constrain the segmentation at these gestational ages. Examples of fetal envelope shape prototypes at different gestational ages (in WA) are provided in Fig. 1, illustrating the great variability of shapes and the need for multiple exemplars in our database.

To encode the variability of the shapes with respect to limb positioning, different positions were generated for fetuses at 13 and 14 WA using the method described in Serrurier et al. (2013), as illustrated in Fig. 2. For younger gestational ages, limbs are not yet sufficiently developed to require such simulations. Eight

¹ The exact resolution is unknown and varies spatially.

² Bradley Smith, University of Michigan (brdsmith@umich.edu) NIH award N01-HD-6-3257 P/G F003637, Imaging performed at the Center for In-Vivo Microscopy, Duke University

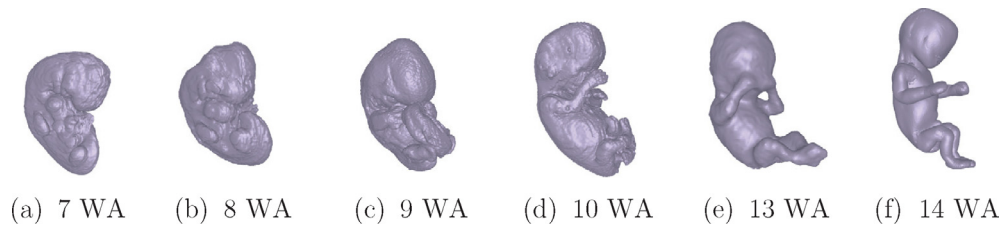


Fig. 1. Examples of fetal envelope shape prototypes at different gestational ages (in WA) used to construct our database.



Fig. 2. Examples of different positions taken by a 14 WA fetus generated with the framework presented in Serrurier et al. (2013) and included in the database. The internal organs visible on this illustration were not used in this work, but only the fetal envelope.

different positions were generated for ages 13 and 14 WA, leading to 30 shapes in our database.

Finally, to take into account the variability with respect to global positioning of the fetus, 14 rotations of the sagittal axis aligned with the CRL, with angles in the range $[-\frac{\pi}{7}, \frac{\pi}{7}]$, were applied to each shape, leading to a final total of 420 shapes in the database.

For each fetus shape model in the database, restricted fetus shape models, consisting of only the trunk and the head (without the limbs), were also generated.

To separate the fetus envelope from the uterus wall, we also constructed a database of fetus back shapes, for each WA, as illustrated in Fig. 3. The back shapes are composed of the posterior part of the fetal torso, using the CRL plane as separator between the posterior and anterior parts. For each WA represented in the database of fetus envelopes, a back shape was selected. These shapes were interpolated to generate models at the missing ages. Linear interpolation was performed along radial directions, radiating from the middle of the crown-rump segment.

2.3. Gray-level statistical distributions & multi-phase level-set framework

2.3.1. Intensity distribution modeling

The intensity distribution modeling was designed to handle a wide range of US images, without prior assumption on the saturation in the low intensity range, and other types of pre-processing

inside the scanning machine. Thus, following the work of Anquez et al. (2013) and Dahdouh et al. (2013), Normal, Rayleigh, Exponential and Gamma distributions are used as possible priors. Nakagami distribution is also used, following (Namburete et al., 2013), to enhance the genericity of the segmentation method.

The mathematical expression of each probability distribution function (pdf) is provided in Table 1, for $I(x)$ representing the image intensity at voxel x .

2.3.2. Choice of the envelope shape model

For a given 3DUS volume to segment, the user is asked to manually determine the CRL by selecting two points: the first one corresponding to the point on the top of the head (the crown) and the second one corresponding to the bottom of the buttocks (the rump).

While CRL is easily defined interactively during a fetal ultrasound examination, its automated measure requires some pre-processing on the recorded 3D US volumes. Indeed, the CRL has to be measured on a sagittal slice (Robinson and Fleming, 1975) of the fetus envelope (head and trunk), while the provided 3D US volumes are usually acquired along arbitrary orientations and slicing of the fetus. Therefore, manual reorientation was performed on each processed 3D US volume to generate a sagittal image slice that includes a suitable anatomical cut of the fetus for accurate CRL measurement (as illustrated in Fig. 4). Some attempts were made to automate this re-orientation step, using a third manually-selected point, but did not lead to a satisfactory level of robustness and accuracy.

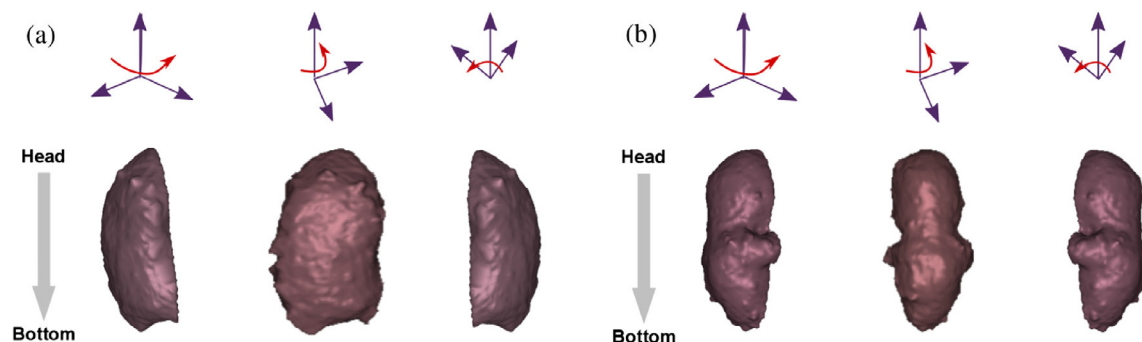


Fig. 3. Fetus back shape models at (a) 10 WA and (b) 13 WA.

Table 1
Formulation of the pdfs of the intensity parametric models used in the level-set segmentation.

Name	pdf $p(I(x))$	Parameters and notations	Image values
Normal	$p(I(x)) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(I(x)-\mu)^2}{2\sigma^2}\right)$	$\mu \in \mathbb{R}, \sigma > 0$	$I(x) \in \mathbb{R}$
Exponential	$p(I(x)) = \lambda e^{-\lambda I(x)}$	$\lambda > 0$	$I(x) \in [0, \infty[$
Rayleigh	$p(I(x)) = \frac{I(x)\exp(-\frac{I(x)^2}{2\sigma^2})}{\sigma^2}$	$\sigma > 0$	$I(x) \in [0, \infty[$
Gamma	$p(I(x)) = I(x)^{\alpha-1} \frac{\beta^\alpha \exp(-\beta I(x))}{\Gamma(\alpha)}$	$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt, \alpha > 0, \beta > 0$	$I(x) \in [0, \infty[$
Nakagami	$p(I(x)) = \frac{2m^m}{\Gamma(m)\Omega^m} I(x)^{2m-1} \exp\left(-\frac{m}{\Omega} I(x)^2\right)$	$\langle \cdot \rangle$ the mean value, $\Omega = \langle I(x)^2 \rangle, m = \frac{\Omega^2}{\langle (I(x)^2 - \Omega)^2 \rangle}$	$I(x) \in [0, \infty[$

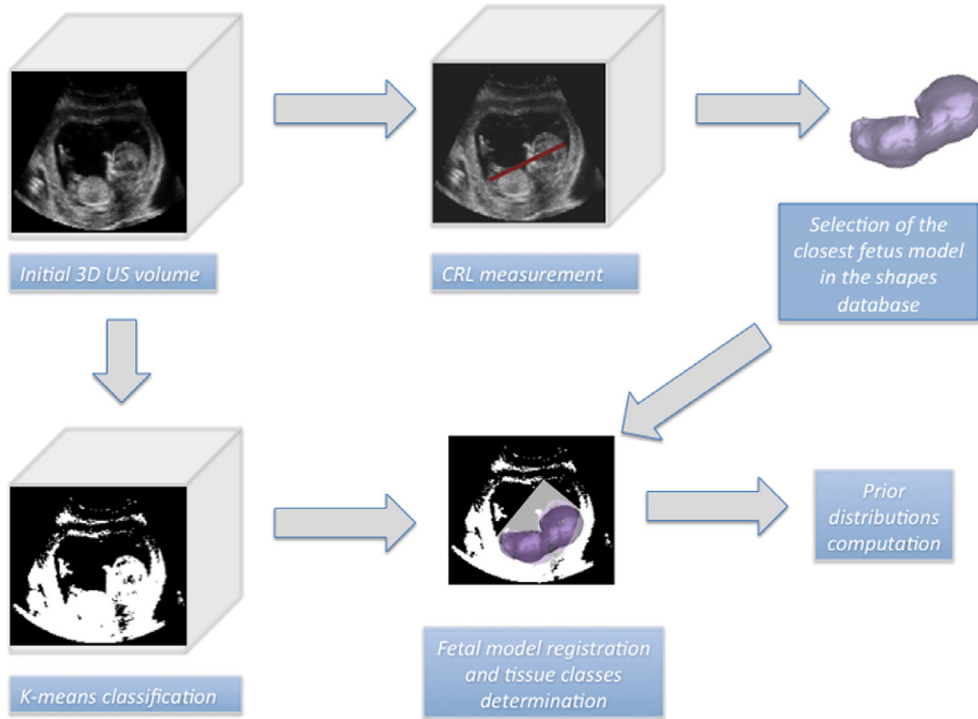


Fig. 4. Pipeline of the initial approximation of masks for the three tissue types to segment: amniotic fluid, fetal envelope and maternal tissues.

Once the pre-registration is performed, the restricted fetus shape model with the closest CRL value is selected and registered on the images via alignment of the CRL lines, as illustrated in Fig. 4. Indeed, during the first trimester of pregnancy, the CRL is one of the most reliable biomarker for determining the fetus age and thus for selecting the closest fetus in terms of global shape according to bio-developmental criteria.

2.3.3. Choice of the intensity distribution model

A k-means classifier is used to partition the 3D US volume into two classes C_{mf} and C_a , approximating respectively the union of fetal and maternal tissues, and the amniotic fluid.

The C_{mf} class is further partitioned into two classes C_m and C_f , corresponding to pixels that are outside and inside the registered restricted fetus model. The C_m class provides a rough approximation of the maternal tissues and C_f of the fetal tissues, without the limbs. The masks of the three classes are then used to identify the intensity pdf that best matches the underlying tissue histogram. Histogram matching is illustrated in Fig. 5, for the maternal and fetal tissues. The matching involves histogram computation, ML estimations of the pdf parameters, removal of the pdf that do not pass the χ^2 test at a level of significance of 0.05, and selection, within the retained candidate pdf, of the one with the best fit measured with the Bayesian information criterion (Schwarz, 1978).

2.3.4. Multi-phase level-set framework

Based on the previous work of Anquez et al. (2013) on two tissue classes, we exploit a multi-phase level set framework initially introduced in Vese and Chan (2002) to segment volumes of images into three tissue classes.

Let Ω be a bounded and open subset of \mathbf{R}^3 and let $I: \Omega \rightarrow \mathbf{R}$ represent the image volume to segment. A given set of n closed surfaces $\{C_1, C_2, \dots, C_n\}$ defines a partition of the image domain in a set of 2^n phases Ω_i . In the case of two level-set functions, we embed these surfaces as the zero-level of signed functions $\{\phi_j\}$, and the image segmentation into four phases is performed by minimizing the following energy:

$$E_{\text{im}} = \sum_{i=1}^4 \int_{\Omega_i} -\log(p_i(I(x), \theta_i)) dx + \sum_{j=1}^2 \nu \int_{\Omega} |\nabla H(\phi_j)| dx \quad (1)$$

where $\nu \geq 0$ is a fixed parameter, H is the Heaviside function computed on each level-set function ϕ_j , and p_i represents the probability of the pixel intensity in phase i so that:

$$p(I(x)) = p_i(I(x), \theta_i), \quad \theta_i \in \Theta_i \quad (2)$$

Since it is well known that the regularization term in Eq. (1) is not able to prevent local segmentation inhomogeneities, especially

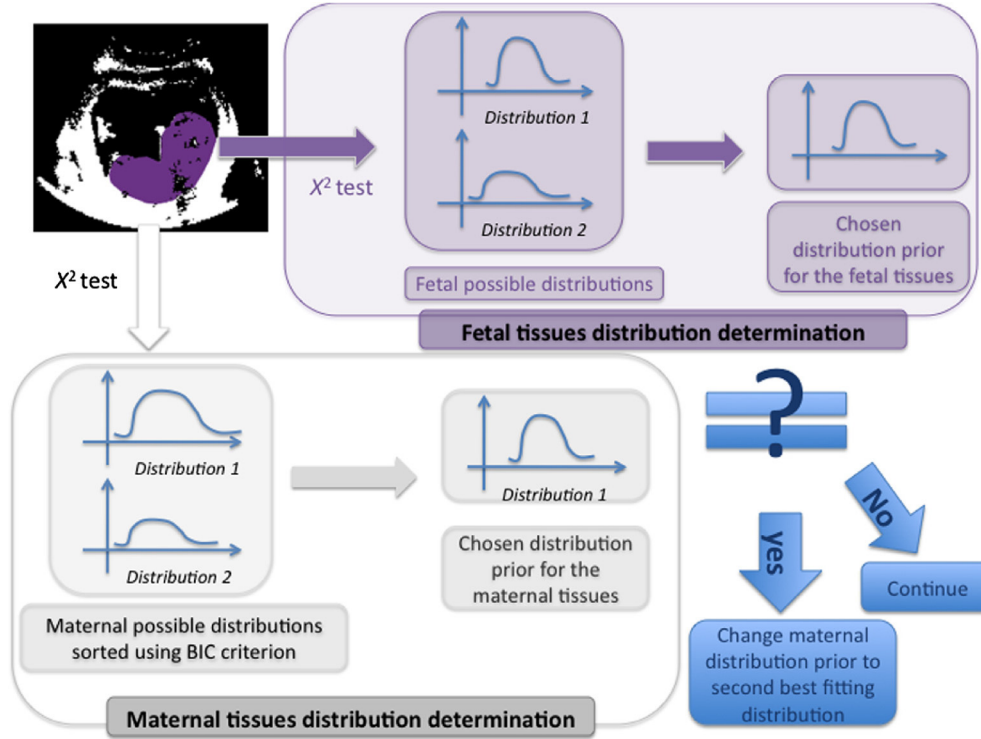


Fig. 5. Illustration of the selection process of the pdf parametric model for the maternal and fetal tissues, to be used in the iterative level-set segmentation process.

on noisy images such as in this work, we rely on an additional shape constraint, described next, to tackle this issue.

2.4. Shape constraint and Legendre moments

Numerous methods have been proposed to add a shape constraint in a level-set based segmentation energy functional. Approaches differ in the encoding of the shape model (for example via PCA (Leventon et al., 2000; Tsai et al., 2003), or Legendre moments (Foulonneau et al., 2009)) and the formulation of the constraint, applied on the zero-level contour or on the whole level-set function. In this work, we reemploy the approach from Wojak et al. (2010) where the shape prior was encoded with Legendre moments and integrated into a variational framework. Legendre moments are indeed efficient to handle the large variability of fetus positions and morphologies illustrated in Section 2.2, while being invariant to scale and translation. The variability, represented within the database of training shapes, is encoded via learning the distribution of a finite set of Legendre moments. This work then introduces a novel shape constraint suited for the multiphase level-set formulation.

2.4.1. Shape encoding with Legendre moments

To use Legendre moment decompositions, the training shapes are first encoded as binary shape images $I_s : [-1, 1]^3 \rightarrow [0, 1]$ where spatial coordinates are normalized in $[-1, 1]$. The $(p + q + r)$ th order 3D Legendre moments of the shape image are defined as:

$$L_{pqr} = \lambda_{pqr} \int_{[-1,1]^3} P_p(x)P_q(y)P_r(z)I_s(x, y, z)dx dy dz \quad (3)$$

with $\lambda_{pqr} = \frac{(2p+1)(2q+1)(2r+1)}{8}$, $(p, q, r) \in \mathbb{N}^3$ and P_i ($i = p, q, r$) the Legendre polynomial defined as:

$$P_i(x) = \sum_{k=0, i-k=\text{even}}^i (-1)^{\frac{i-k}{2}} \frac{1}{2^i} \frac{(i+k)!x^k}{\left(\frac{i-k}{2}\right)! \left(\frac{i+k}{2}\right)! k!} \quad (4)$$

Working with a finite number N of moments (set to 60 in all our experiments) for each dimension, an estimate of I_s is given by:

$$\tilde{I}_s(x, y, z) = \sum_{p=0}^N \sum_{q=0}^p \sum_{r=0}^q \lambda_{p-q, q-r, r} P_{p-q}(x)P_{q-r}(y)P_r(z) \quad (5)$$

The final shape is obtained by thresholding the reconstructed shape image (we denote by T_{LM} the thresholding value). The computation of the 3D Legendre moments can be performed using the fast method proposed by Hosny (2011). Translation is achieved by reformulating the Legendre moments in Eq. (3) by replacing (x, y, z) by $(x - x_0, y - y_0, z - z_0)$ where (x_0, y_0, z_0) are the coordinates of the center of mass of the shape.

Using the CRL measurement used for the pixels intensity distribution computation, the 3D transformation in terms of scale, translation and rotation between the currently studied fetus and the database shapes is computed. This transformation is used to achieve scale and rotation invariance.

2.4.2. Formulation of the shape constraint

To constrain the segmentation process using a shape prior, all the training shapes are decomposed into N Legendre moments. Since for each gestational age, different positions of the limbs exist in the database, we do not assume an equiprobability between all training shapes. Indeed, while in Dahdouh et al. (2013) this assumption was made to counterbalance the small size of the shape database, we drive here the segmentation process toward an *a priori* shape by minimizing the distance between the currently studied fetus and the shapes in the database in terms of age, while making no assumption on the fetal position.

A weight on the training shapes, depending on the CRL, is thus introduced and the new energy term is formulated as follows:

$$E = E_{im} + E_p(L_s) \quad (6)$$

where the shape prior term $E_p(L_s)$ is computed on the Legendre moments L_s of the phase s to constrain. Indeed, since we want to

control only the shape of the fetus envelope, we constrain only the phase Ω_s containing it. The shape prior energy term is expressed as:

$$E_p(L_i) = \begin{cases} 0 & \text{if } i \neq s \\ \frac{-1}{2(1+\lambda)} \left(L_i - \sum_{k=1}^{N_{\text{ref}}} w(d_k) L_k^{\text{ref}} \right)^2 & \text{if } i = s \end{cases} \quad (7)$$

with N_{ref} the number of prototype shapes, L_i the vector of Legendre moments $\{L_{pqr}\}$ of the characteristic function χ_i of the considered phase i , L_k^{ref} the vector of Legendre moments of the k th reference shape, λ the weight to be given to the shape prior and $w(d_k^i)$ the weight to be given to the k -th shape with respect to the phase i and the shape encoded by L_i .

The parameter λ is used to weight the influence of the data fidelity and the shape constraint terms. The data fidelity term tends to aggregate maternal and fetal tissues together while the shape constraint tends to propose a shape corresponding to a linear combination of the shapes in the prior database. The choice of this value will be further discussed in Section 2.5.

The weights $w(d_k)$ are computed using the CRL segments, described in Section 2.3.2, as a mean to objectively encode the age difference between the fetus being segmented and the k -th prototype shape. The weights are defined as:

$$w(d_k) = \begin{cases} 1 & \text{if } d_k = 0 \\ \frac{N_{\text{eq}} + 1}{\alpha} & \text{if } d_k > 0 \end{cases} \quad (8)$$

where $d_k = |\text{CRL}_k - \text{CRL}_f|$ measures the difference of CRL values between the k -th shape in the database (CRL_k) and the segmented fetus (CRL_f), N_{eq} is the number of elements in the shape database for which $d_k = 0$ (i.e. same CRL-based gestational age as the fetus being segmented) and α is a scalar value used to normalize the sum of the weights of the N_{ref} shapes in the database to: $\sum_{k=1}^{N_{\text{ref}}} w(d_k) = 1$. Using $N_{\text{eq}} + 1$ avoids a division by 0. Fig. 6 illustrates the values of the weights $w(d_k^i)$ for the segmentation of a 12 WA fetus with $N_{\text{ref}} = 420$ shapes in the database.

We note here that, since all the training shapes are represented by means of their Legendre moments and their CRL measurement, there is no limitation on the proposed method regarding scalability. Indeed, when introducing new shape models, it is sufficient to compute their Legendre moments and provide their crown rump length measurement. Eq. (8) can then be computed, by taking these new shapes into account. These steps have a linear complexity in the number of shapes. All the other steps of the method can be applied directly, without need for further training, nor added complexity.

2.4.3. Iterative level-set segmentation process

Since we want to control only the shape of the fetus envelope, we only constrain the phase containing it. The other phases are evolved based only on their gray level statistical appearance models and the regularization prior detailed in Eq. (1), as explained in the optimization procedure detailed next.

The fetal tissue phase is detected according to its statistics so that it matches the fetal tissue prior distribution. After the level-set functions have evolved over few iterations, the algorithm can estimate the parameters of the intensity distributions of the different phases. The shape prior is then added in the evolution procedure to constraint the phase representing the fetal envelope. Indeed, since the level-set functions are initialized using a set of cylinders regularly interleaved over the image, it is not possible to detect the fetal phase over the first iterations since the phases are not homogeneous and the intensity distribution statistics are therefore not representative. At each iteration, the data fidelity term using shape statistics is computed to prevent the addition of structures not belonging to the currently seg-

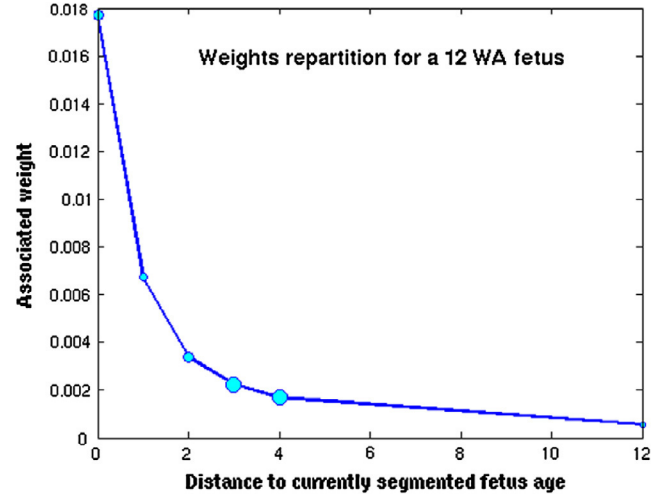


Fig. 6. Distribution of shape prior weights $w(d_k^i)$ for each shape in our database to segment a 12 WA fetus. The circle size represents the number of shapes that belong to individual age stages.

mented fetus. Details on the initialization and refinement procedures are given in the algorithm described in Appendix A.

Considering an artificial time t , Euler Lagrange equations of ϕ_j evolution are written as:

$$\frac{\partial \phi_1}{\partial t} = \delta_\epsilon(\phi_1) \left\{ \nu \text{div} \left(\frac{\nabla \phi_1}{|\nabla \phi_1|} \right) - [\log(p_1(I(x), \theta_1) - \log(p_2(I(x), \theta_2))H(\theta_2)) \right. \\ \left. + [\log(p_3(I(x), \theta_3) - \log(p_4(I(x), \theta_4))(1 - H(\theta_2))) \right\} \quad (9)$$

$$\frac{\partial \phi_2}{\partial t} = \delta_\epsilon(\phi_2) \left\{ \nu \text{div} \left(\frac{\nabla \phi_2}{|\nabla \phi_2|} \right) - [\log(p_1(I(x), \theta_1) - \log(p_2(I(x), \theta_2))H(\theta_1)) \right. \\ \left. + [\log(p_3(I(x), \theta_3) - \log(p_4(I(x), \theta_4))(1 - H(\theta_1))) \right\} \quad (10)$$

with δ_ϵ the regularized version of the Dirac delta function as defined in Vese and Chan (2002).

2.5. Parameterization and initialization of the segmentation process

2.5.1. Level-sets initialization and reinitialization

Following a standard paradigm, as in Angelini et al. (2006), we initialized the level set functions with two set of cylinders regularly interleaved over the image. During the first set of iterations, for the rough approximation of the tissue phases, we did not reinitialize the level set functions. During the refinement phase, level set functions are recomputed at each step from the zero-level defined on the shape returned by the moments reconstruction.

2.5.2. Parameterization

Time step: The time step Δt was set to 0.1 for all our experiments.

Choice of ν : ν was set to 100 in all our experiments.

Choice of T_{LM} , the reconstruction threshold: Using a linear combination of Legendre moments does not lead to a binary shape image, as seen in Fig. 7, and a thresholding operation is therefore needed to generate the final fetal envelope.

Few works have discussed specifically the choice of this threshold value, since most of the approaches using Legendre encoding of shapes do not use the reconstruction procedure. In Zhang et al. (2011) or Dahdouh et al. (2013) the threshold was arbitrarily set to 0.5. In this work, we decided not to use a pre-fixed threshold value, since the same threshold could lead to under- or over-estimation of the fetal envelope, depending on the case. Instead,

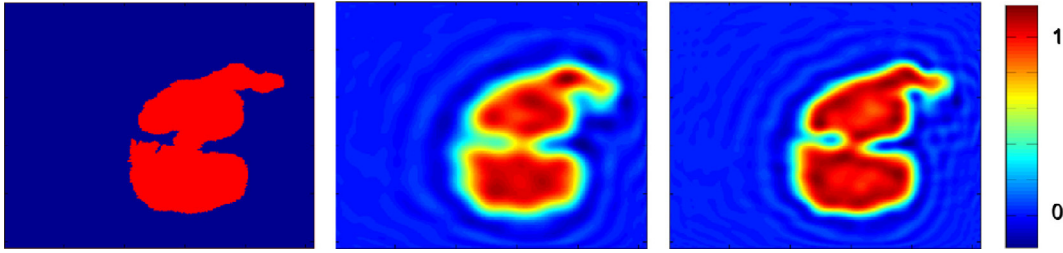


Fig. 7. Binary shape (left) and its approximation using 40 (center) and 60 (right) Legendre moments, leading to non-binary Legendre-moment shape images.

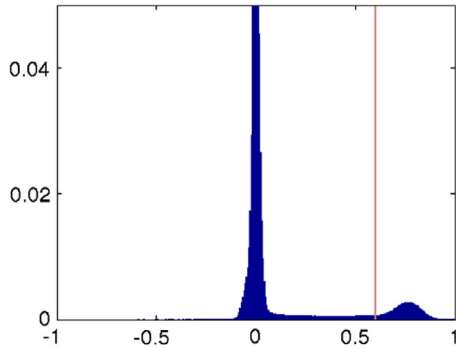


Fig. 8. Histogram of the normalized Legendre-moment shape image of a fetal envelope and illustration of the selection of the threshold value T_{LM} as the left tail of the second mode.

an empirical case-specific selection of the threshold value is performed, based on the observed histogram of the normalized reconstructed shape image. As illustrated in Fig. 8, the histogram of the normalized shape images is bimodal: the first mode represents the background and the second one the reconstruction of the fetal envelope. The threshold value was set at the left side of the second peak (defined as the center of the bin before the peak, for a 255 bin histogram). The segmentation is first evolved with the threshold T_{LM} computed at each iteration, for 5 iterations, and then with a fixed value (set to the value at iteration 5) for the remaining iterations until convergence.

Number of moments. To enable the statistical prior to encode fine shape details, a high number of moments is needed, as illustrated in Fig. 9 and in accordance with the work of Foulonneau et al. (2009).

To empirically set the required number of Legendre moments, we measured Dice, sensitivity and specificity indices on 30 cases from 7 to 22 WA, comparing the approximated shape to the ground truth, for different numbers of moments. We also measured CPU computation time. Results of this study are reported in Fig. 10. Using this set of reconstructions as well as Fig. 9, we concluded that the best tradeoff between computational time and quality was obtained for a number of moments equal to 60.

Choice of λ value: The parameter λ was set empirically to 0.5 for all our experiments. This value was found to be a good tradeoff be-

tween the image data fidelity term and the shape constraint on the fetal envelope.

Convergence criteria: Convergence of the level set segmentation was measured via the following criterion on the fetal envelope shape: five consecutive iterations with no significant changes of the fetal shape, measuring the differences, between two consecutive iterations, in the position of the contour and in the number of voxels belonging to the fetal envelope. It has to be equal to 0 for convergence to be reached. The convergence was reached in at most 200 iterations for all the experiments in the study, with a mean value of 100 iterations per case.

2.6. Post-processing of the fetus envelope

A comparison of the level-set segmentation results obtained with the algorithm previously introduced, without and with the shape constraint, is shown in Fig. 11. As seen in this figure, using the shape prior greatly improves the fetus envelope segmentation, but some maternal tissues can still remain attached to the fetus along the lower uterus boundaries, in cases where the fetal back is in contact. To address this issue, a dedicated post-processing step is now introduced, exploiting the fetus back models illustrated in Fig. 3.

As illustrated on the 13 WA fetus in Fig. 11, a post-processing step is needed to remove maternal tissues on the back of the fetal envelope, when it is in contact with the uterus wall. The role of this post-processing step is to separate the front of the fetus from its back, to clean the back of the fetus and then to re-attach both parts, as detailed next. Fig. 12 details the main steps of the post-processing procedure.

The actual fetus age, estimated previously from the CRL, is first used to select the appropriate back model in the database. This back model Ω_{back} is then rigidly registered with the segmented fetus χ^{final} using the CRL measurement as done in Section 2.3.2.

The final fetal phase χ_f^{res} is obtained using the morphological operation:

$$\chi_f^{res} = I_b \cup I_f \quad (11)$$

with I_f the front of the final fetus, and I_b the back of the final fetus, defined as:

$$I_f = cc_{MAX}(\chi^{final} \setminus \Omega_{back}) \quad (12)$$

$$I_b = \chi^{final} \cap \Omega_{back} \quad (13)$$



Fig. 9. Illustration of the effect of the number of Legendre moments used to reconstruct a fetal envelope shape on one case at 7 WA. The initial shape corresponds to a manually traced contour.

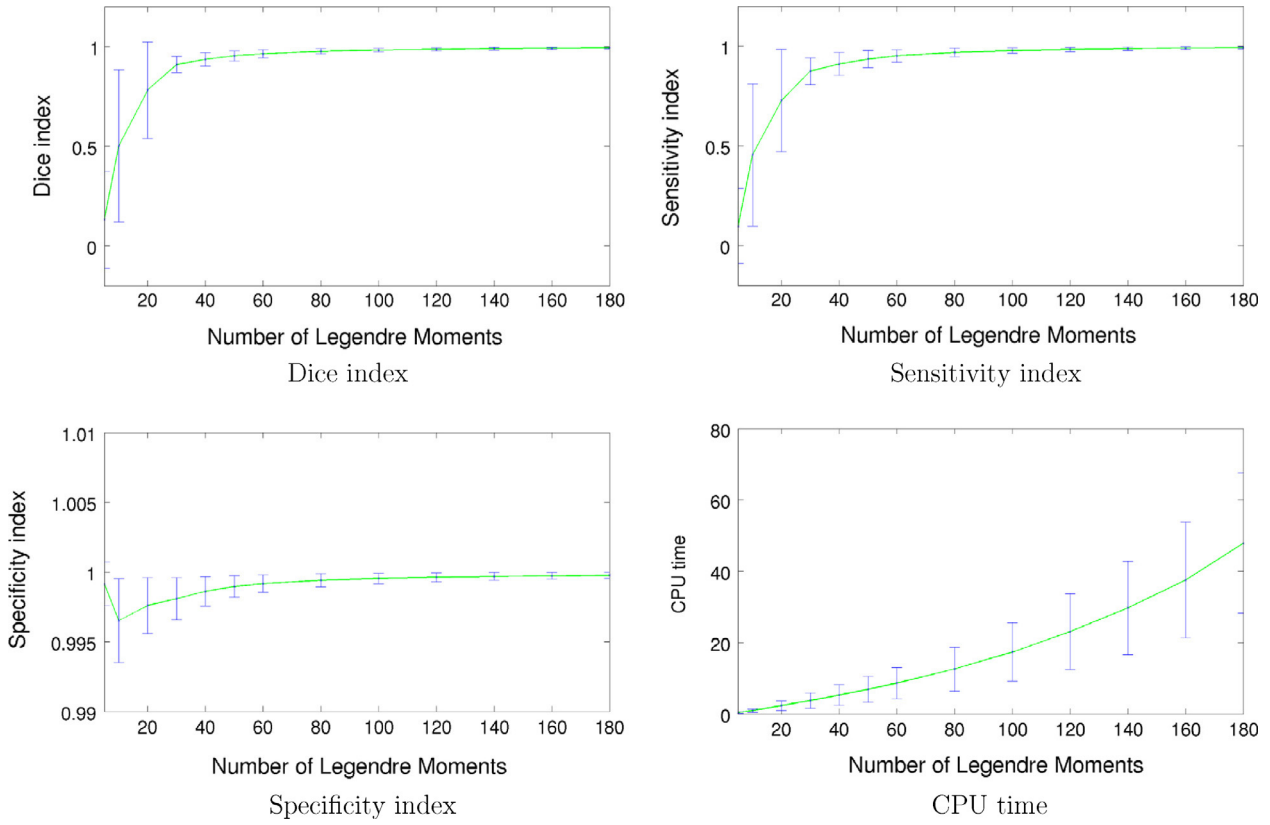


Fig. 10. CPU computation time (in seconds) and Dice, sensitivity and specificity indices for fetal envelope shapes reconstructed with increasing numbers of Legendre moments, comparing to manual tracing. Average index values are reported, for 30 cases covering different gestational ages.

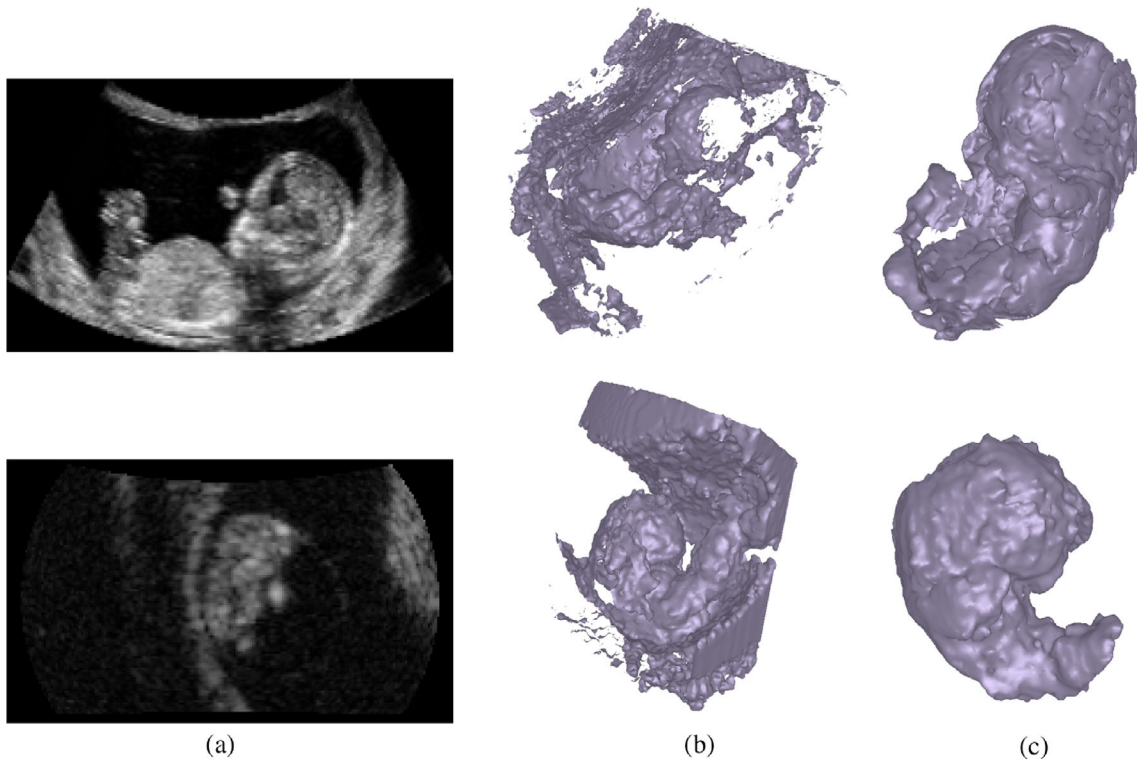
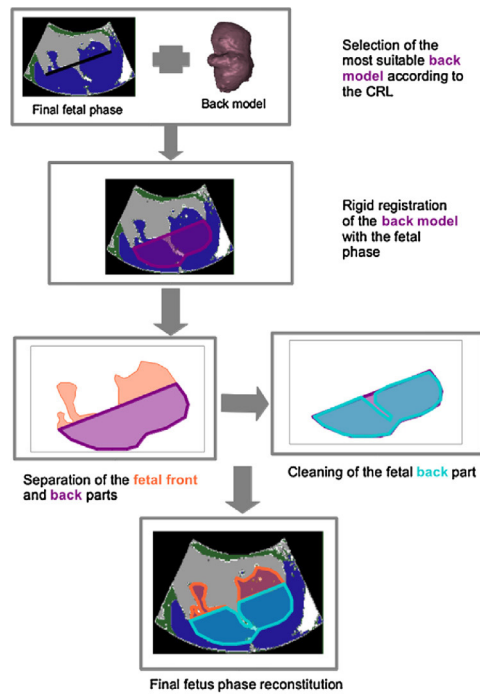
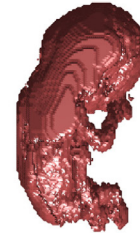


Fig. 11. Segmentation of two fetal envelopes with the proposed method at 13 (top line) and 8 (bottom line) WA. (Column a) Original slices from the 3D US volumes. (Column b) Fetal envelopes segmented without the shape prior. (Column c) Fetal envelopes segmented with the shape prior.



(a)



(b)

(b) Effects of the post-processing to remove maternal tissues attached to the segmented fetal envelope of a 13 WA fetus. Subplots (a) represent the fetal envelope before, and (b) after post-processing.

(a) Post processing of the fetal envelope segmentation exploiting the back models.

Fig. 12. Post processing of the fetal envelope segmentation.

cc_{MAX} being the largest connected component retrieval operation. $\chi^{final} \setminus \Omega_{back}$ removes all but the largest segmented fetus envelope components connected to the back model Ω_{back} , erasing the few remaining tissue components still attached to the back of the fetus. Thus, the largest remaining connected component is assumed to be the front of the fetus envelope. Otherwise, the amount of remaining maternal tissues is too important for the segmentation procedure to

be considered successful. Non-fetal tissues are efficiently removed by this post-processing, as illustrated in Fig. 12, while the umbilical cord still remains attached to the fetus in this case.

3. Results

To evaluate the performances of our method, a set of 14 3D US volumes of images was segmented and the extracted fetal

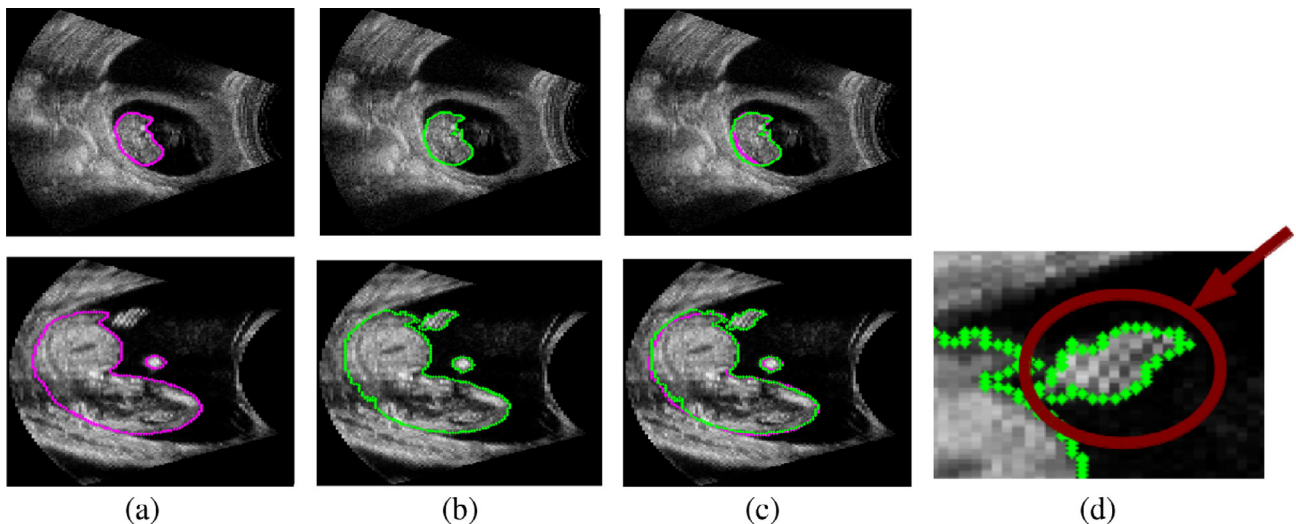


Fig. 13. Manual contours (a) and automated segmentation (b) overlaid on the original 3DUS images of two fetal volumes acquired at 10 (1st row) and 13 (2nd row) WA. (c) Superposition of both automated and manual contours on the original images. (d) Zoom on the umbilical cord for the 2nd volume, which was segmented as being part of the fetal envelope automatically but not manually.

envelopes were compared to manually traced envelopes provided by an expert, using the Dice and similarity indices, and the specificity and sensitivity measures (mathematical definitions are provided in Appendix C).

The volume with the 22 WA fetus was also segmented, to test the robustness of the segmentation framework with respect to transducer type, older age of the fetus and handling of a truncated field of view.

3.1. Quantitative measures

We obtained the following average levels of Dice, similarity, sensitivity and specificity: 0.8, 0.7, 0.75 and 0.98.

If we restrict the evaluation to the same dataset as the one used in Dahdouh et al. (2013), average similarity, sensitivity and specificity values are: 0.75, 0.89 and 0.98 while they were of 0.72, 0.89 and 0.97 in Dahdouh et al. (2013). Since the two sets of measures do not follow a Normal distribution, we performed a Wilcoxon signed rank test for paired data to determine whether the difference between the two studies in term of similarity is statistically significant or not. The p -value being equal to 0.01, the equality hypothesis was rejected at the 5% significance level and the differences between the two studies can be considered significant. Some segmentation results are visually illustrated in Fig. 13.

The proposed formulation was therefore able to improve the segmentation performance, and solves certain limitations such as the remaining attachment of some maternal tissues to the fetal envelope as well as a better handling, even if it is still not perfect, of the various possible fetal positions.

3.2. Inter-expert segmentation study

We further evaluated the quality of the segmentation results by measuring inter-expert variability of the manual segmentations. Three different experts were asked to segment three volumes of images at 9, 12 and 13 WA. The third author, being an experienced gynecologist with more than 20 years of practice was chosen as the reference manual tracer. Dice and similarity measures were computed as reported in Table 2. As we can see, due to the high level of noise in the images and the complexity of the shape of the fetus envelope, inter-expert variability is quite high. The number of tested segmentations is too low to conclude on the statistical significance of the observed variability. We would like to emphasize the difficulty of manually segmenting these fetal images due to strong heterogeneities inside the body, dropouts of the US signal along structures parallel to the beam and a complex morphology of the contours to trace with multiple disconnect components corresponding to different body parts sliced in 2D. For all these reasons, enforcing 3D spatial continuity of the contours is also extremely tedious. Hence, this manual tracing task is radically different from tracing a left-ventricular cavity or a fetal skull for example. Overall the inter-expert similarity range is [0.68, 0.85] and the Dice range is [0.80, 0.92].

Similarity and Dice values obtained between the proposed automated method and the manual ground-truth fall within this range.

Table 2

Inter-expert variability of manual segmentation of the fetal envelope on 3D US images. The contours provided by expert #3 were used as the ground-truth to evaluate the segmentation method.

Similarity index				Dice index			
Volume	Expert 1	Expert 2	Expert 3	Volume	Expert 1	Expert 2	Expert 3
Volume 1	0.83	0.85	1	Volume 1	0.9	0.92	1
Volume 2	0.68	0.78	1	Volume 2	0.80	0.87	1
Volume 3	0.83	0.80	1	Volume 3	0.91	0.89	1

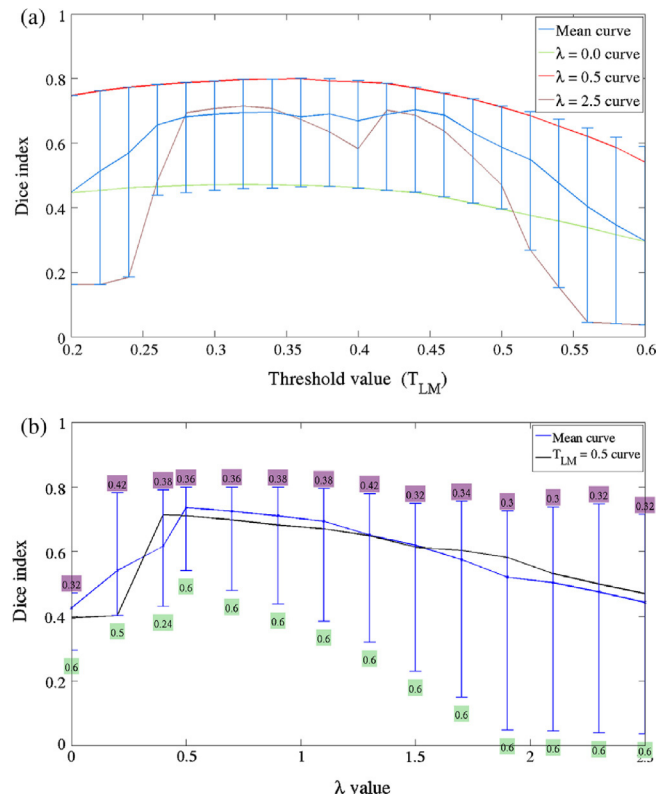


Fig. 14. (a) Evolution of the Dice index with respect to the T_{LM} parameter for three different λ parameter values, and the average of Dice index values over all tested λ values. (b) Evolution of the Dice index with respect to the λ parameter value for $T_{LM} = 0.5$ and the average of Dice index values over all tested T_{LM} values. Numbers in the green (resp. purple) squares report the T_{LM} parameter value returning the worst (resp. best) Dice index for each λ . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.3. Robustness to parameters variation

To evaluate the robustness of our method, we studied the influence of three different parameters of the method. First, the manual selection of the two points defining the CRL value was studied. We observed that the final segmentation results are not affected by a variation of up to 5 pixels of the position of the CR segment extremities. Variations of more than 5 pixels can lead to an erroneous estimation of the fetal age and choice of different fetal shape models, itself leading to a decrease in the segmentation quality. We then studied the influence of both λ , parameter used in Eq. (7) to weight the shape prior energy term, and T_{LM} , the threshold applied on the reconstructed shape image, using an experimental plan described next. The T_{LM} value can be seen as the minimal probability under which a pixel is considered as not belonging to the fetal shape. For each volume of images in the database, the segmentation result (before post-processing) was computed for the couple (λ, T_{LM}) with $\lambda \in [0.0, 2.5]$ and $T_{LM} \in [0.2, 0.6]$. All possible combinations were tested, with increments of 0.2 on the λ values

and of 0.02 on the T_{LM} values. The Dice index was computed for each segmentation result.

Fig. 14(a) reports the evolution of the Dice index as a function of T_{LM} for different λ values for one volume of images (results were similar for the other volumes). The blue curve represents the mean Dice value obtained for all the tested λ values for a given T_{LM} . The red curve represents the results obtained for $\lambda = 0.5$ while the green and purple one respectively represent the results obtained for $\lambda = 0.0$ and $\lambda = 2.5$. As we can see, the use of $\lambda = 0.5$ allows us to obtain near optimal results for each tested T_{LM} on this volume.

In the same way, Fig. 14(b) reports the evolution of the Dice index as a function of λ for different T_{LM} values for one image. The blue curve represents the mean Dice value obtained for all the tested T_{LM} values for a given λ . The black curve represents the Dice index values obtained with a T_{LM} value set to 0.5 while numbers in the green (respectively purple) squares represent the T_{LM} value leading to the worst (respectively best) segmentation result for a given λ .

As we can see, the results obtained with $T_{LM} = 0.5$, while close to the mean Dice result, never lead to the best segmentation. Moreover, depending on the λ value, the optimal threshold value varies, thus justifying the proposed automated case-specific threshold selection.

A joint optimization of the couple of parameters (λ, T_{LM}) could lead to even better segmentation results but was not investigated in this work, given the already high Dice indices obtained with the proposed approach and the added computational cost that it would generate.

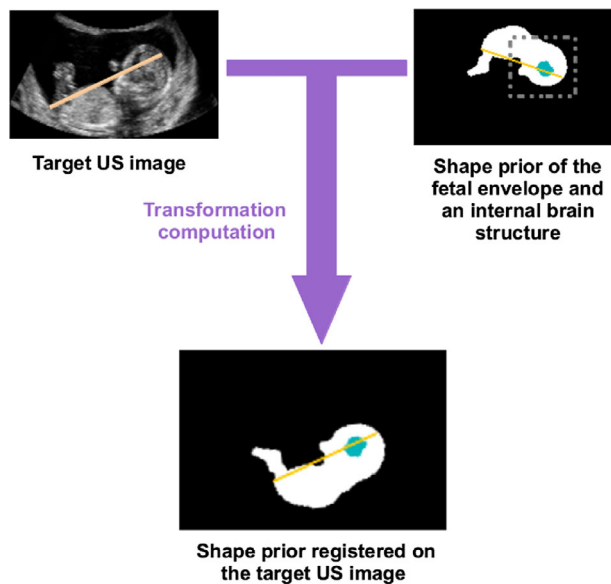


Fig. 15. Registration of the fetal shape prior with an internal brain structure. The CRL is used to register both the envelope shape prior and the internal brain structure. The bounding box around the skull is used to limit the segmentation process inside a reasonable region around the actual brain structures.

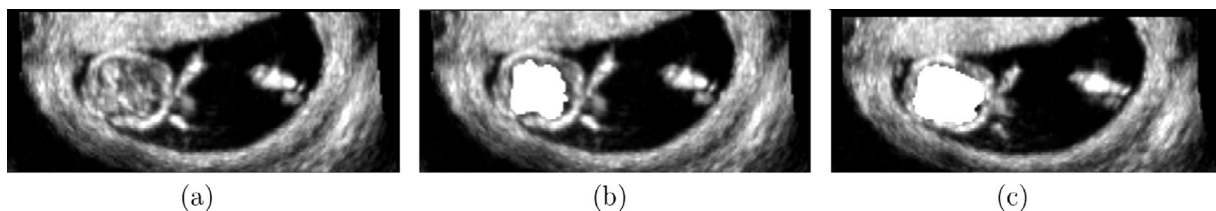


Fig. 16. (a) Original 3D US volume. (b) Manual segmentation of the choroid plexus. (c) Automatic segmentation using the proposed framework (similarity, sensitivity and specificity are 0.78, 0.83 and 0.99).

4. Conclusion and discussion

4.1. Conclusion

A shape-constrained multi-phase level-set segmentation framework was introduced to segment the fetal envelope on 3D US images. The 3D segmentation method embeds statistical priors of pixel intensity distributions in the different parts of the utero-fetal unit and a shape prior for the fetus envelope. Shape information is encoded with translation-invariant Legendre moments and shape priors were learned over a variety of cases and gestational ages. Initialization of the segmentation process is performed semi-automatically by asking the user to determine the crown-rump segment and its length (CRL). The CRL value is used to compute weighting factors associated with shape priors. Indeed, the fetus age is estimated using this measure, and used to weight shape priors depending on their distance, in terms of age, to the segmented fetus. A post-processing step is also proposed, to remove, from the segmented fetus envelope, any attached maternal tissues using a generic back model of the fetus at a given fetal age. While not always necessary, this step significantly improved segmentation results in cases where the sagittal slices were not aligned with the sagittal orientation of the fetus trunk (e.g. when the fetus was laying on his side). Tests on 14 clinical cases provided satisfying results when visually and quantitatively compared to manual segmentations. Sensitivity of the segmentation method with respect to shape-prior weighting parameters and CRL input measure was also studied and has confirmed the robustness of the parameterization paradigms used in the proposed segmentation framework. To test the genericity of our method, the segmentation framework presented in this paper was extended to further segment, within the fetal envelope, internal structures such as the choroid plexuses. This adaptation of the previously presented segmentation algorithm has shown promising results, as explained next.

Extension of the segmentation to additional structures. During the registration of the fetus shape priors described in Section 2.3.2, a set of 4 choroid plexuses shape priors, generated on cases from 12 to 13 WA, was also registered, within a bounding box defined around the skull, as shown in Fig. 15. The bounding box is used to restrain the search space and deformations of the shape priors inside the brain. Once this initialization is done, the same multiphase level-set segmentation framework is applied, combining both shape and intensity priors (except for the post-processing step). Parameters were computed as explained previously.

Preliminary results on a case at 13 WA are illustrated in Fig. 16. Similarity, sensitivity and specificity were 0.78, 0.83 and 0.99 for this example.

4.2. Discussion

In the proposed segmentation method, some issues remain, such as the need to detach the umbilical cord from the fetus envelope and the need to compensate for incomplete fetal segmentation. Regarding the umbilical cord, it is currently segmented with the

fetal tissues phase as shown in the second row of Fig. 13. This problem also occurs with manual segmentations and specific removal of the umbilical cord has often to be performed due to the difficulty to differentiate it from the limbs. Regarding the incompleteness of the fetal envelope segmentations, there are two types of remaining artefacts: (1) limbs, especially legs and feet, are sometimes missing in the final segmentations and (2) the fetus envelope tends to be overall smaller than the manually segmented one.

These issues are mainly due to not enough different limb positions being represented in our database. Indeed, we observed that when increasing the size of the database by including more limb positions, the first type of artefact decreased drastically.

Concerning the second problem, it mainly comes from the fact that all the fetal shapes used as priors represent fetuses in a flexed position so errors occur when we try to segment a shape in a more extended position.

The use of manual segmentations as a ground truth in this type of segmentation task is to be used with caution, as emphasized by the prospective study on the variability of the experts contours. Indeed, the fetal envelope is a complex three dimensional structure, with multiple parts, and US images are very noisy, requiring great care and skill while tracing on 2D slices, and making the manual segmentation task very time consuming. Thus, beyond the difficulty to visualize the boundaries of the structures on US images, the time needed to segment a whole fetus is important enough to lead to errors due to eyestrain and tiredness. It is therefore not surprising to observe very high inter-expert variability even when using the same tracing software tool.

Most studies done on 2D fetal US image data aim at computing biometric measures such as the bi-parietal diameter, the abdominal circumference or the femur length. These measures are mainly computed during the second and third trimester of pregnancy using geometrical primitives such as ellipsoids and line segments. While Dice indices reported in these fetal biometrical studies, such as (Ciurte et al., 2012) or (Rueda et al., 2014), are typically around 0.95 for the fetal head segmentation, in our task, inter-expert agreement only reached 0.92 with a lowest score at 0.8. A most relevant comparison would be with the 2D fetal femur segmentation task whose main challenge is the presence of big homogeneities inside the femur (Rueda et al., 2014). In Rueda et al. (2014), two methods are compared and present respectively an average Dice value of 0.77 and 0.74. When compared to these values, the average Dice value of 0.8 obtained for our segmentation task appears to be satisfactory.

Very close to this work, in Gupta et al. (2011), authors combined conditional random fields with wavelet-based textural features and support vector machine classification to segment the fetal envelope on 2D fetal ultrasound images. However, no quantitative error measures are reported and it is therefore difficult to compare our results to theirs.

At an early stage of pregnancy, we can stipulate that the computation of the fetus volume enables to date it more finely than simple 2D morphometric measures currently performed in routine exams. Clinical validation of this stipulation could be performed in the future on a set of pregnancies from *in vitro* fertilization. At a later stage of pregnancy, the direct measure of the fetal volume could enable to monitor more precisely the fetal growth. Finally, our proposed segmentation involves shape modeling and shape matching to a “normal” template, which could pave the way to automated malformation detection.

Possible improvements of the proposed method include the use of a 3D pose normalization instead of using our initial registration, the adaptation of the priors weights throughout the segmentation procedure in order to promote the most suitable limbs positions and the automated co-optimization of both λ and T_{LM} parameters.

Future work also includes the extension of the proposed approach to other fetal structures.

Acknowledgment

This work has been partly supported by a grant from ANR-JST within the FETUS project. The authors would like to thank Cochin, Beaujon and Saint Vincent de Paul hospitals for providing data as well as medical expertise, and B. Smith from the University of Michigan for providing 3D embryo models. We also would like to thank Joe Wiart and his team at Orange Labs for their collaboration.

Appendix A. Level-set evolution algorithm

{Initialization}

For few iterations

1. Initialize $\{\phi_j\}_{j=1,2}$ as $\{\phi_j^0\}_{j=1,2}$
2. Evolution of $\{\phi_j\}_{j=1,2}$ by minimizing Eq. (1) using the Euler–Lagrange derivatives (Eqs. (9) and (10)) with respect to the level set functions and the distribution parameters, leading to the following updates between iterations n and $n + 1$:

$$\{\phi_j^n\}_{j=1,2} \rightarrow \{\phi_j^{n+1}\}_{j=1,2} \quad (\text{A.1})$$

$$\{\phi_j^{n+1}\}_{j=1,2} \rightarrow \{\chi_i^{n+1}\}_{i=1,4} \quad (\text{A.2})$$

$$\{\theta_i^n\}_{i=1,4} \rightarrow \{\theta_i^{n+1}\}_{i=1,4} \quad (\text{A.3})$$

with θ_i the distribution parameters of phase i .

End

{Optimization procedure}

Do

1. Identification of the phases and the characteristic functions, denoted χ_f , χ_m and χ_l that include respectively the fetal tissues, the maternal tissues and the amniotic fluid, based on the pixel intensity distributions. A χ^2 test is used to determined the phases corresponding to the different tissues.
2. Evolution of $\{\phi_j\}_{j=1,2}$ by minimizing Eq. (1) using the Euler–Lagrange derivatives with respect to the level set functions and the distribution parameters as in step (2) above.
3. Computation of translation and scale invariant Legendre moments for χ_f :

$$\chi_f^{n+1} \rightarrow L_{pqr} \quad (\text{A.4})$$

4. Update of L_{pqr} via gradient descent:

$$L'_{pqr} = L_{pqr} + \frac{\partial E_{\text{prior}}}{\partial L_{pqr}} \quad (\text{A.5})$$

5. Estimation of the fetal envelope mask Ω_f from the updated Legendre moments, using T_{LM} as the reconstruction threshold:

$$L'_{pqr} \rightarrow \Omega_f \quad (\text{A.6})$$

6. Shape refinement using pixel intensity distributions:

(a) Extract fetal tissues pixels shared by Ω_f and χ_f^{n+1} :

$$\Omega_f \cap \chi_f^{n+1} \rightarrow \Omega_f^* \quad (\text{A.7})$$

(b) Test the candidate pixels provided by the shape constraint and coming from the maternal tissues to be added to the fetal tissue phase based on a statistical test function:

$$\Omega_{mr} = \emptyset$$

For each connected component $cc_k \in \Omega_f \setminus \Omega_f^*$

If $cc_k \in \chi_m^{n+1}$
 then If $\text{TEST}_{\theta_f^n}(I, cc_k, \Omega_f^*)$
 then

$$\begin{aligned}\Omega_f^* &= \Omega_f^* \cup cc_k \\ \Omega_{mr} &= \Omega_{mr} \cup cc_k\end{aligned}$$

end
 end
 end

(c) Update the characteristic functions of the fetal and maternal tissue phases:

$$\begin{aligned}\chi_f^{n+1} &= \Omega_f^* \\ \chi_m^{n+1} &= \chi_m^{n+1} \setminus \Omega_{mr}\end{aligned}$$

with cc_k denoting connected components and $\text{TEST}_{\theta_f^n}(I, cc_k, \Omega_f^*)$ a χ^2 statistical test used to decide whether the intensities I of pixels in cc_k can be considered to belong to the statistical distribution predicted by the pdf model of χ_f with parameters θ_f^n updated on Ω_f^* .

7. Update intensity distribution parameters for the fetal and maternal tissue phases, based on χ_f^{n+1} and χ_m^{n+1} : assign each connected component removed from the fetal phase to the most probable phase according to their statistics. Knowing χ_f^{n+1} , update $\{\chi_i^{n+1}\}_{i=1,4}$

$$\{\chi_i^{n+1}\}_{i=1,4} \rightarrow \{\phi_j^{n+1}\}_{j=1,2} \quad (\text{A.8})$$

Details of the involved procedure are given in [Appendix B](#).

Until the fetal tissue phase converges during p consecutive iterations. In our experiments, p was set to 5 iterations. Convergence is reached when the sum of differences in term of pixels between a shape at iteration n and a shape at iteration $n + 1$ is null.

Appendix B. ϕ computation from phases χ

Let $\{\phi_j\}_{j=1,2}$ denote the two level-sets, $\{\phi_j^-\}_{j=1,2}$ and $\{\phi_j^+\}_{j=1,2}$ respectively their negative and positive parts, and $\{\chi_i^{n+1}\}_{i=1,4}$ the four phases:

If we define $\{\chi_i^{n+1}\}_{i=1,4}$ as:

$$\begin{aligned}\chi_1 &= \phi_1^- \cap \phi_2^- \\ \chi_2 &= \phi_1^- \cap \phi_2^+ \\ \chi_3 &= \phi_1^+ \cap \phi_2^- \\ \chi_4 &= \phi_1^+ \cap \phi_2^+\end{aligned}$$

then:

$$\begin{aligned}\chi_1 \cup \chi_2 &= (\phi_1^- \cap \phi_2^-) \cup (\phi_1^- \cap \phi_2^+) \\ \chi_1 \cup \chi_3 &= (\phi_1^- \cap \phi_2^-) \cup (\phi_1^+ \cap \phi_2^-)\end{aligned}$$

and therefore:

$$\begin{aligned}\chi_1 \cup \chi_2 &= \phi_1^- \\ \chi_1 \cup \chi_3 &= \phi_2^-\end{aligned}$$

from which $\{\phi_j\}_{j=1,2}$ can be retrieved.

Appendix C. Definition of the evaluation measures for segmentation quality

To compare two contours represented as binary images $\text{Im}_{\text{result}}$ and $\text{Im}_{\text{reference}}$ (used as the reference), we used the following measures:

Similarity index

$$\frac{|\text{Im}_{\text{result}} \cap \text{Im}_{\text{reference}}|}{|\text{Im}_{\text{result}} \cup \text{Im}_{\text{reference}}|} \quad (\text{C.1})$$

Dice index

$$\frac{2|\text{Im}_{\text{result}} \cap \text{Im}_{\text{reference}}|}{|\text{Im}_{\text{result}}| + |\text{Im}_{\text{reference}}|} \quad (\text{C.2})$$

Sensitivity index

$$\frac{\text{TP}}{\text{TP} + \text{FN}} \quad (\text{C.3})$$

Specificity ratio

$$\frac{\text{TN}}{\text{TN} + \text{FP}} \quad (\text{C.4})$$

with: TP, the number of true positives, TN the number of true negatives, FP the number of false positives and FN the number of false negatives.

References

- Abdallah, Y., Daemen, A., Guha, S., Syed, S., Naji, O., Pexsters, A., Kirk, E., Stalder, C., Gould, D., Ahmed, S., Bottomley, C., Timmerman, D., Bourne, T., 2011. Gestational sac and embryonic growth are not useful as criteria to define miscarriage: a multi-center observational study. *Ultrasound Obst. Gynecol.* 38, 503–509.
- Angelini, E.D., Song, T., Laine, A., 2006. Homogeneity measures for multiphase level set segmentation of brain MRI. In: *IEEE International Symposium on Biomedical Imaging: From Nano to Macro ISBI'06*, pp. 746–749.
- Anquez, J., Angelini, E., Grange, G., Bloch, I., 2013. Automatic segmentation of antenatal 3-d ultrasound images. *IEEE Trans. Biomed. Eng.* 60, 1388–1400.
- Aviram, R., Shpan, D., Markovitch, O., Fishman, A., Tepper, R., 2004. Three-dimensional first trimester fetal volumetry: comparison with crown rump length. *Early Hum. Dev.* 80, 1–5.
- Blaas, H.G.K., Taipale, P., Torp, H., Eik-Nes, S.H., 2006. Three-dimensional ultrasound volume calculations of human embryos and young fetuses: a study on the volumetry of compound structures and its reproducibility. *Ultrasound Obstet. Gynecol.* 27, 640–646.
- Bottomley, C., Bourne, T., 2009. Dating and growth in the first trimester. *Best Pract. Res. Clin. Obst. Gynaecol.* 23, 439–452.
- Bottomley, C., Daemen, A., Mukri, F., Papageorghiou, A., Kirk, E., Pexsters, A., De Moor, B., Timmerman, D., Bourne, T., 2009. Functional linear discriminant analysis: a new longitudinal approach to the assessment of embryonic growth. *Hum. Reprod.* 24, 278–283.
- Campbell, S., Warsof, S.L., Little, D., Cooper, D.J., 1985. Routine ultrasound screening for the prediction of gestational age. *Obst. Gynecology* 65, 613–620.
- 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.
- Ciurte, A., Rueda, S., Bresson, X., Nedevschi, S., Papageorghiou, A., Noble, J., 2012. Ultrasound image segmentation of the fetal abdomen: a semi supervised patch-based approach. In: *MICCAI workshop on Perinatal and Paediatric imaging: PaPI*.
- Dahdouh, S., Serrurier, A., Grange, G., Angelini, E.D., Bloch, I., 2013. Segmentation of fetal envelope from 3D ultrasound images based on pixel intensity statistical distribution and shape priors. In: *IEEE International Symposium on Biomedical Imaging: From Nano to Macro ISBI'13*, pp. 1014–1017.
- Destrempe, F., Meunier, J., Giroux, M.F., Soulez, G., Cloutier, G., 2009. Segmentation in ultrasonic B-mode images of healthy carotid arteries using mixtures of Nakagami distributions and stochastic optimization. *IEEE Trans. Med. Imaging* 28, 215–229.
- Eltoft, T., 2006. Modeling the amplitude statistics of ultrasonic images. *IEEE Trans. Med. Imaging* 25, 229–240.
- Falcon, O., Peralta, C.F.A., Cavoretto, P., Auer, M., Nicolaidis, K.H., 2005. Fetal trunk and head volume in chromosomally abnormal fetuses at 11 + 0 to 13 + 6 weeks of gestation. *Ultrasound Obstet. Gynecol.* 26, 517–520.
- Foulonneau, A., Charbonnier, P., Heitz, F., 2009. Multi-reference shape priors for active contours. *Int. J. Comput. Vision* 81, 68–81.
- Gupta, L., Sisodia, R., Pallavi, V., Firtion, C., Ramachandran, G., 2011. Segmentation of 2D fetal ultrasound images by exploiting context information using conditional random fields. In: *EMBC, 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 7219–7222.
- Hosny, K.M., 2011. Fast and low-complexity method for exact computation of 3D Legendre moments. *Pattern Recogn. Lett.* 32, 1305–1314.
- Jardim, S.M.G.V.B., Figueiredo, M.A.T., 2005. Segmentation of fetal ultrasound images. *Ultrasound Med. Biol.* 31, 243–250.

- Levaillant, J., 2006. Intérêt de l'échographie 3D-4D en échographie foetale et gynécologique: principes et indications. *J. Radiol.* 87, 1969–1992.
- Leventon, M.E., Grimson, W.E.L., Faugeras, O., 2000. Statistical shape influence in geodesic active contours. In: Conference on Computer Vision and Pattern Recognition, pp. 316–323.
- Martins, W.P., Ferriani, R.A., Nastri, C.O., Filho, F.M., 2008. First trimester fetal volume and crown-rump length: comparison between singletons and twins conceived by in vitro fertilization. *Ultrasound Med. Biol.* 34, 1360–1364.
- Martins, W.P., Nastri, C.O., Barra, D.A., Navarro, P.A., Filho, F.M., Ferriani, R.A., 2009. Fetal volume and crown-rump length from 7 to 10 weeks of gestational age in singletons and twins. *Eur. J. Obst. Gynecol. Reprod. Biol.* 145, 32–35.
- Namburete, A., Rahmatullah, B., Noble, J., 2013. Nakagami-based AdaBoost learning framework for detection of anatomical landmarks in 2D fetal neurosonograms. *Ann. Br. Mach. Vision Assoc. (BMVA)* 2, 1–16.
- Noble, A., Boukerroui, D., 2006. Ultrasound image segmentation: a survey. *IEEE Trans. Med. Imaging* 25, 987–1010.
- Paragios, N., Jolly, M.P., Taron, M., Ramaraj, R., 2005. Active shape models and segmentation of the left ventricle in echocardiography. In: Kimmel, R., Sochen, N.A., Weickert, J. (Eds.), *Scale-Space*, pp. 131–142.
- Reljic, M., 2001. The significance of crown rump length measurement for predicting adverse pregnancy outcome of threatened abortion. *Ultrasound Obstet. Gynecol.* 17, 510–512.
- Roberts, N., Thilaganathan, B., 2007. The role of ultrasound in obstetrics. *Obst., Gynecol. Reprod. Med.* 17, 79–85.
- Robinson, H.P., Fleming, J.E.E., 1975. A critical evaluation of sonar crown-rump length measurements. *BJOG: Int. J. Obst. Gynaecol.* 82, 702–710.
- Rousian, M., Koning, A., Oppenraaij, R.v., Hop, W., Verwoerd-Dikkeboom, C., Spek, P.V.d., Exalto, N., Steegers, E., 2010. An innovative virtual reality technique for automated human embryonic volume measurements. *Hum. Reprod.* 25, 2210–2216.
- Rueda, S., Fathima, S., Knight, C., Yaqub, M., Papageorgiou, A., Rahmatullah, B., Foi, A., Maggioni, M., Pepe, A., Tohka, J., Stebbing, R., McManigle, J., Ciurte, A., Bresson, X., Cuadra, M., Sun, C., Ponomarev, G., Gelfand, M., Kazanov, M., 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.
- Sarti, A., Corsi, C., Mazzini, E., Lamberti, C., 2005. Maximum likelihood segmentation of ultrasound images with Rayleigh distribution. *IEEE Trans. Ultrason., Ferroelectr. Freq. Control* 52, 947–960.
- Schwarz, G.E., 1978. Estimating the dimension of a model. *Ann. Stat.* 2, 461–464.
- Serrurier, A., Dahdouh, S., Captier, G., Calmels, V., Adamsbaum, C., Bloch, I., 2013. 3D articulated growth model of the fetus skeleton, envelope and soft tissues. *Innov. Res. Biomed. Eng.* 34, 349–356.
- Shankar, P.M., 1995. A model for ultrasonic scattering from tissues based on the k distribution. *Phys. Med. Biol.* 40, 1633–1649.
- Smeets, N.A., Winkens, B., Prudon, M., van de Ven, J., Gondrie, V.P., van Deursen, F., Oei, S.G., 2013. The predictive value of first trimester fetal volume measurements, a prospective cohort study. *Early Hum. Dev.* 89, 321–326.
- Sonek, J., 2007. First trimester ultrasonography in screening and detection of fetal anomalies. *Am. J. Med. Genet.* 145C, 45–61.
- Tsai, A., Yezzi, A., Wells, W., Tempany, C., Tucker, D., Fan, A., Grimson, W., Willsky, A., 2003. A shape-based approach to the segmentation of medical imagery using level sets. *IEEE Trans. Med. Imaging* 22, 137–154.
- Tsui, P.H., Hsu, C.W., Ho, M.C., Chen, Y.S., Lin, J.J., Chang, C.C., Chu, C.C., 2010. Three-dimensional ultrasonic Nakagami imaging for tissue characterization. *Phys. Med. Biol.* 55, 5849–5866.
- Vese, L.A., Chan, T.F., 2002. A multiphase level set framework for image segmentation using the Mumford and Shah model. *Int. J. Comput. Vision* 50, 271–293.
- Wojak, J., Angelini, E.D., Bloch, I., 2010. Introducing shape constraint via Legendre moments in a variational framework for cardiac segmentation on non-contrast CT images. In: *International Conference on Computer Vision Theory and Applications*, pp. 209–214.
- 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.
- Zhang, Y., Matuszewski, B.J., Histace, A., Precioso, F., 2011. Statistical shape model of Legendre moments with active contour evolution for shape detection and segmentation. In: *Computer Analysis of Images and Patterns*, vol. 6854, pp. 51–58.