Semantic-based Indexing of Fetal Anatomies from 3-D Ultrasound Data Using Global/Semi-local Context and Sequential Sampling

Gustavo Carneiro[1], Fernando Amat[2], Bogdan Georgescu[1], Sara Good[3], Dorin Comaniciu[1]
[1] Siemens Corporate Research, Integrated Data Systems Department, Princeton, NJ
[2]Stanford University, Department of Electrical Engineering, Stanford, CA
[3]Siemens Medical Solutions, Innovations Division, Mountain View, CA

Abstract

The use of 3-D ultrasound data has several advantages over 2-D ultrasound for fetal biometric measurements, such as considerable decrease in the examination time, possibility of post-exam data processing by experts and the ability to produce 2-D views of the fetal anatomies in orientations that cannot be seen in common 2-D ultrasound exams. However, the search for standardized planes and the precise localization of fetal anatomies in ultrasound volumes are hard and time consuming processes even for expert physicians and sonographers. The relative low resolution in ultrasound volumes, small size of fetus anatomies and inter-volume position, orientation and size variability make this localization problem even more challenging. In order to make the plane search and fetal anatomy localization problems completely automatic, we introduce a novel principled probabilistic model that combines discriminative and generative classifiers with contextual information and sequential sampling. We implement a system based on this model, where the user queries consist of semantic keywords that represent anatomical structures of interest. After queried, the system automatically displays standardized planes and produces biometric measurements of the fetal anatomies. Experimental results on a held-out test set show that the automatic measurements are within the inter-user variability of expert users. It resolves for position, orientation and size of three different anatomies in less than 10 seconds in a dual-core computer running at 1.7 GHz[1].

1. Introduction

The foremost physicians and researchers in the fields of obstetrics and gynecology (OBGYN) have advocated that the use of 3-D ultrasound (3DUS) data for diagnosis and regular exams is potentially one of the next breakthroughs in radiology [1, 4]. Fetal biometric measurements is one of the main OBGYN applications and represent an important factor for high quality obstetrics health care. These measurements are used for estimating the gestational age (GA) of the fetus, assessing of fetal size and monitoring of fetal growth and health. Nowadays, these measurements require the manual search for the standardized plane using 2-D ultrasound (2DUS) images, which is a cumbersome activity that contributes to the excessive length in clinical obstetric examinations [1], leading to serious repetitive stress injuries (RSI) for the sonographers [11, 14] and more expensive health care. Compared to 2DUS, the main advantages of 3DUS are the following: 1) substantial decrease in the examination time [1], 2) possibility of post-exam data processing without requesting additional visits of the patient [1], 3) the ability of experts to produce 2-D views of the fetal anatomies in orientations that cannot be seen in common 2-D ultrasound exams [3, 12] and 4) potential reduction of RSI for sonographers.

One of the main obstacles for the widespread use of 3DUS is the requirement of extensive manipulation on the part of the physician or the sonographer in order to reach standard planes so that the fetal biometric measurements can be performed. The learning curve to understand these manipulation steps is quite large even for expert users [1]. Usually, expert users need to find several landmarks in order to reach the sought anatomy. For example, the standardized plane for measuring the lateral ventricles in the fetus brain is referred to as the transventricular plane (Fig. 1), and the user must search for the cavum septi pellucidi, frontal horn, atrium, and choroids plexus in order to reach this plane. Since the fetus is oriented in an arbitrary position in each volume, an expert sonographer may require several minutes to localize all the necessary structures in a basic examination[9]. Therefore, the use of 3DUS for fetal biometrics measurements depends heavily on the ability of expert users to navigate in ultrasound volumes of fetuses. This ability can be significantly improved if certain fetal anatomies could be automatically indexed through the use of semantic keywords (e.g., cerebellum or lateral ventricles). The indexing of these anatomies involves the display of the standard plane and the biometric measurement of the fetal anatomy (Fig. 2).

[1]Fernando Amat completed this work while he was with the Integrated Data Systems Department at Siemens Corporate Research

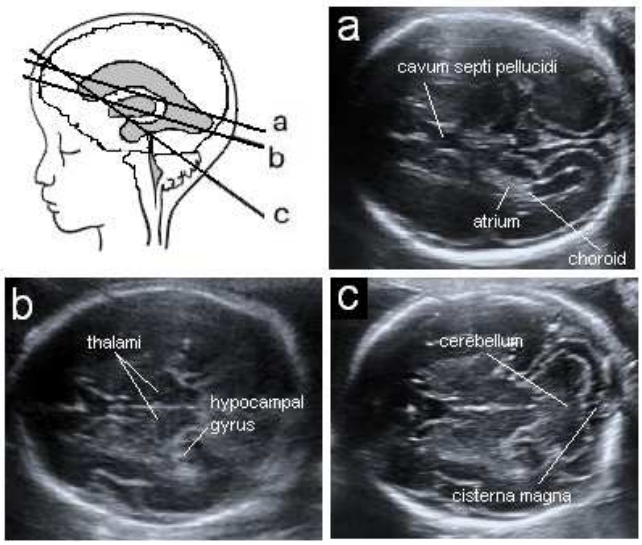


Figure 1. Axial views of the fetal head [12]. (a) Transventricular plane; (b) transthalamic plane; (c) transcerebellar plane.

In this paper, we propose a system that receives as inputs an ultrasound volume containing the head of a fetus and a semantic keyword as a user query, and automatically displays the standardized plane for measuring the requested anatomy along with its biometric measurement. We focus on the automatic indexing of the following three fetal anatomies: cerebellum (CER), cisterna magna (CM), and lateral ventricles (LV). The biometric measurements of these anatomies are performed according to the guidelines of the International Society of Ultrasound in Obstetrics and Gynecology [12].

Three dimensional ultrasound imaging presents many challenges to develop a reliable clinical system to automatically index such fetus structures. The first challenge is the quality of the images. Shadows, speckle noise and other artifacts create a low contrast image with blurry edges (see Fig. 2). Even if it is easy for the human eye to navigate and recognize structures it is difficult to adapt common feature extraction techniques to 3DUS datasets. Second, the size of the anatomical structures depends on the age of the fetus, which brings a high variance in the model for each structure. Finally, the position and orientation of each fetus is completely arbitrary, making impossible to constraint the search space in 3D.

To the best of our knowledge, we are not aware of similar methods for automatic semantic-based indexing of anatomical structures in 3DUS. Most of the literature in the field of 3DUS is confined to the problems of (semi-)automatic segmentation of specific anatomical structures [5, 17, 6] and registration [16, 18]. It is unclear whether such methods can work for the system introduced in this paper because we do not face neither a segmentation nor a registration problem. In computer vision there are methods for recognizing 3D objects using range images [7], but these applications are different in the sense that the system basically works with surfaces instead of actual volumes, so a direct comparison with these methods is not possible. There has been similar works to the one presented in this paper using 3-D magnetic resonance imaging (3DMRI) data. For example, Tu et al. [21] proposed a combination of discriminant classifier based on the probabilistic boosting tree (PBT) [20] for appearance and generative classifier based on principal components analysis (PCA) for shape, where the weights for these two terms are learned automatically. This is applied to the segmentation of eight brain structures, where the system takes eight minutes to run. Recently, Zheng et al. [24] introduced a new segmentation of heart structures using 3-D computed tomography (3DCT) based on discriminant classifiers and marginal space learning [10]. This system achieves the segmentation of four heart structures in less than eight seconds. Despite the similarities with the 3DMRI and 3DCT systems above, the challenges for our system in 3DUS are different. For instance, instead of precise anatomy segmentation we aim for a precise biometric measurement, involving correct pose estimation of the anatomy. Also, imaging characteristics of 3DMRI and 3DCT are different from that of 3DUS making the extension of MRI and CT applications not straightforward. Finally, the orientation of anatomical structures in 3DMRI and 3DCT is generally better constrained than that of 3DUS.

In order to provide a completely automatic solution for the problem of fetal anatomy indexing in 3DUS, we introduce a *novel principled probabilistic model* that combines *discriminative and generative* classifiers with *contextual information and sequential sampling*. We take advantage of over 200 hundred annotated ultrasound volumes for three anatomical structures: cerebellum, cisterna magna and lateral ventricles. This large number of annotations allows us to use PBT [20] to learn relevant features over a large pool of 3-D Haar features [13, 23, 22] and steerable features[24]. Both features have been shown in the literature to be efficiently computed and to be effective as a feature space for boosting classifiers. The pose estimation for the three fetal anatomies mentioned above involves 3-D position, orientation and scale resulting in 7 degrees of freedom (i.e., a total of 21 degrees of freedom for all anatomies) in a typical volume of dimensions 250x200x150 voxels. This large dimensional search space makes a brute force approach not practical. Thus, to make the problem tractable, we use sequential sampling [24, 10] and contextual information [19].

The basic idea of sequential sampling is as follows: the initial parameter space is partioned into sub-spaces of increasing dimensionality, where the PBT classifiers are trained *sequentially* in each of these sub-spaces using bootstrap samples. We use the same sequence during the detection process. This approach allows for significant gains in terms of detection and training time complexities as we shall see in Sec. 2.3.

Contextual information is based on two cues: 1) global context based on the detection of the *fetal skull*; and 2) semi-

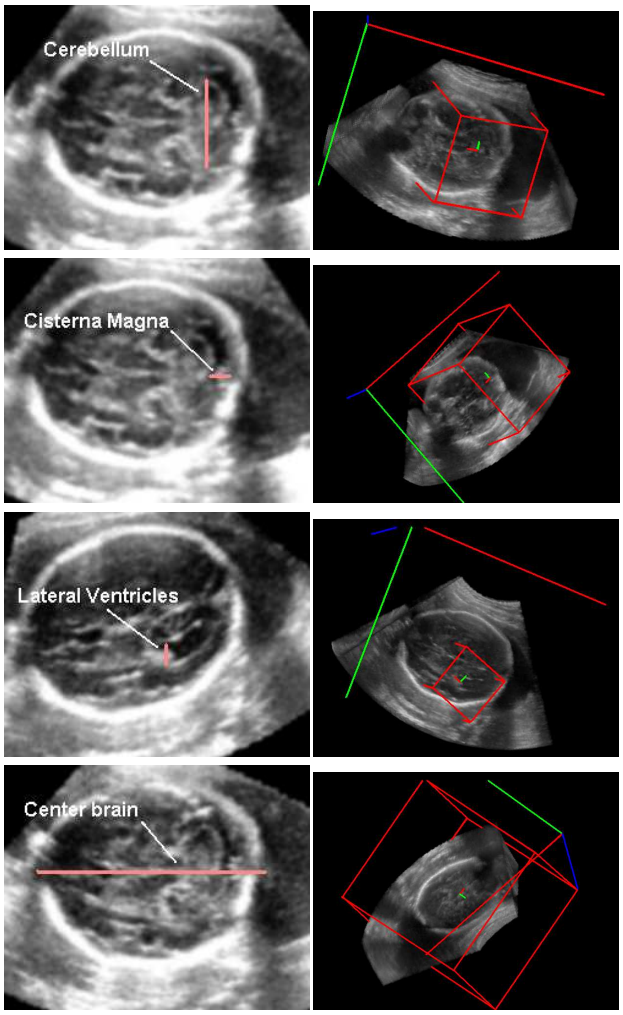


Figure 2. Biometric measurements (first column) and respective 3-D sample (second column) of Cerebellum (row 1), Cisterna Magna (row 2), Lateral Ventricles (row 3), and center of the brain (row 4). Each 3-D sample is represented by a box with position, scale and orientation.

local context based on the relative position, orientation and scale between anatomies (Sec. 2.2). These cues are modeled with generative classifiers. The fetal skull is the largest and most visible structure in a 3D ultrasound. Thus, it can be used as a reference to constrain the search space of all other anatomies in the brain. Sec. 2.3 shows how this constrain makes the system extremely scalable. The addition of new fetal brain anatomies shall not have a severe impact on the search complexity of the approach.

This fully automatic approach performs fast[2] (under 10 seconds in a dual-core PC at 1.7GHz) and robustly. It locates position, orientation and size of all the three anatomies with an error similar to the inter-user variability, allowing physicians and sonographers to quickly navigate through ultrasound volumes of fetal heads. The current method can be seen as a first step to create a large scale semantic-based fetal structure retrieval from 3DUS, where the users type a semantic keyword and the system returns the structure in the volume making the 3D navigation much easier and faster.

In the remaining of the paper we first present the details of our approach in Sec. 2. We then present the training protocol in Sec. 3, and experimental results in Sec. 4. We conclude the work in Sec. 5.

2. Automatic Measurement of Fetal Anatomy

The input for our system is an ultrasound volume containing the head of a fetus between 13 to 35 weeks of age. After 35 weeks, the ultrasound signal has difficulty penetrating the fetal skull. The user may query the system using a limited vocabulary of semantic keywords. Each keyword represents an anatomy of interest that the user wants to visualize and measure. In particular we consider the following three anatomies: cerebellum, cisterna magna, and lateral ventricles (Fig. 2). Once the user selects the keyword, the system automatically shows the standard plane of visualization and the respective biometric measure.

2.1. Problem Definition: a Probabilistic Framework

A volume is a 3-D mapping $V : \mathbb{R}^3 \rightarrow [0, 255]$. A sub-volume containing a particular anatomical structure is represented by a vector containing position, size and orientation, as follows:

$$\theta_s = [\mathbf{p}, \sigma, \mathbf{q}] \in \mathbb{R}^{\mathbf{7}}, \tag{1}$$

where $\mathbf{p} = [\mathbf{x}, \mathbf{y}, \mathbf{z}] \in \mathbb{R}^{\mathbf{3}}$ is the three dimensional center of the sub-volume, $\sigma \in \mathbb{R}$ represents its size, $\mathbf{q} = [\mathbf{q_1}, \mathbf{q_2}, \mathbf{q_3}] \in \mathbb{R}^{\mathbf{3}}$ represents orientation[3] and s represents a specific anatomy (here, $s \in \{CB, CER, CM, LV\}$, where CB stands for center brain. See Fig. 2). The main goal of the system is to determine the sub-volume parameters of all the anatomies of interest:

$$[\theta^*_{CER}, \theta^*_{CM}, \theta^*_{LV}] = \underset{\theta_{CER}, \theta_{CM}, \theta_{LV}}{\operatorname{argmax}} P(\theta_{CER}, \theta_{CM}, \theta_{LV} | V) \tag{2}$$

where $P(\theta_{CER}, \theta_{CM}, \theta_{LV}|V)$ indicates a probability measure of the anatomy parameters given the volume V. The search space for this case is $O\left(\left(M^7\right)^L\right) = O\left(M^{21}\right)$, where we assume that each dimension can be partioned into M values, and $L = 3$ is the number of anatomies to detect. Typical value for M is in the order of 100, which makes (2) intractable. There has been strong indications that context is interesting for pruning the search space, which has the potential to improve the accuracy and increase the speed of recognition systems [15, 19]. Here we adopt two different types of context information: global and semi-local contexts. The global context is provided by the center of the

[2] We anticipate that a fully optimized code will run under three seconds.

[3] Orientation is represented using quaternions. See subsection 2.4

brain (CB) structures that is derived from the whole skull of the fetus (Fig. 2). CB is the largest and most distinctive feature in a 3D fetal ultrasound, so it can be found reliably in most datasets, and consequently, it can constrain the search space for the other anatomies. Thus, Eq. 2 can be denoted as

$$[\theta^*_{CER}, \theta^*_{CM}, \theta^*_{LV}] = \underset{\theta_{CER},\theta_{CM},\theta_{LV}}{\operatorname{argmax}} \int_{\theta_{CB}} P(\theta_{CB}, \theta_{CER}, \theta_{CM}, \theta_{LV}|V) d\theta_{CB}. \tag{3}$$

Subsection 2.2 explains how the semi-local context constrains even more the search space of the sought anatomy given the anatomies already found.

Assuming the existence of the random variable $y_s = \{-1, 1\}$ for $s \in \{CB, CER, CM, LV\}$, where $y_s = 1$ indicates the presence of the anatomy s, we have:

$$P(\theta_{CB}, \theta_{CER}, \theta_{CM}, \theta_{LV}|V) = P(\{y_s = 1\}_{s\in\{CB,CER,LV,CM\}}|\theta_{CB}, \theta_{CER}, \theta_{CM}, \theta_{LV}, V) \tag{4}$$

We train discriminative classifiers that are capable of computing actual posterior probabilities (e.g., PBT [20]) for each anatomy, so the following probabilities can be computed: $P(y_s = 1|\theta_s, V)$ for $s \in \{CB, CER, CM, LV\}$ (hereafter, we denote $P(y_s = 1|\theta_s, V) = P(y_s|\theta_s, V)$). Using the Bayes rule, (4) can be derived to:

$$P(y_{LV}|y_{CB}, y_{CER}, y_{CM}, \theta_{CB}, \theta_{CER}, \theta_{CM}, \theta_{LV}, V) \cdot P(y_{CB}, y_{CER}, y_{CM}|\theta_{CB}, \theta_{CER}, \theta_{CM}, \theta_{LV}, V)$$

which can be further derived to:

$$P(y_{LV}|\theta_{LV}, V) \cdot P(y_{CB}, y_{CER}, y_{CM}|\theta_{CB}, \theta_{CER}, \theta_{CM}, V) \cdot \frac{P(\theta_{LV}|y_{CB}, y_{CER}, y_{CM}, \theta_{CB}, \theta_{CER}, \theta_{CM}, V)}{P(\theta_{LV}|\theta_{CB}, \theta_{CER}, \theta_{CM}, V)}$$

Notice that we assume that the probability of the presence of LV based on the feature values depends only on θ_{LV} and V, but the probability distribution of θ_{LV} depends on the detection and parameters of other anatomies. This is a common assumption of parts independence but geometry dependency [2]. Also, we assume that the conditional distribution of θ_{LV} given all other parameters is a uniform distribution because there is no notion about the actual presence of the other anatomies. Finally, (4) can be written as follows:

$$\begin{aligned} P(\theta_{CB}, \theta_{CER}, \theta_{CM}, \theta_{LV}|V) = & \\ & P(y_{LV}|\theta_{LV}, V) P(y_{CM}|\theta_{CM}, V) P(y_{CER}|\theta_{CER}, V) \\ & P(y_{CB}|\theta_{CB}, V) P(\theta_{CER}|y_{CB}, \theta_{CB}, V) \cdot \\ & P(\theta_{CM}|y_{CB}, y_{CER}, \theta_{CB}, \theta_{CER}, V) \\ & P(\theta_{LV}|y_{CB}, y_{CER}, y_{CM}, \theta_{CB}, \theta_{CER}, \theta_{CM}, V), \end{aligned} \tag{5}$$

where the first four terms are the posterior probabilities of each anatomy, and the remaining terms account for the global and semi-local context. The detection probability described in (5) suggests a sequential detection where CB is detected first, followed by CER, then CM, and finally LV[4]. This means that the complexity of the detection was reduced from $O\left(M^{7L}\right)$ in its original form (2) to $O\left((L+1) \times M^7\right)$.

2.2. Semi-local Context

The basic idea underlying semi-local context is to, during the detection process, use the parameter values of the detected anatomies to estimate a distribution in the parameter space for the subsequent anatomies to be detected. The use of semi-local context has been recently exploited [2], but here we generalize this approach to 3-D environments. From (1), we see that there are position, scale, and orientation parameters, but let us first focus on how to estimate the position parameters. It is possible to determine an orthonormal matrix $\mathbf{R_s} \in \mathbb{R}^{\mathbf{3\times 3}}$ with the three axis of the coordinate system lying in its rows using the orientation parameters (see for example the axis for the boxes in Fig. 2). In order to produce scale invariant estimates of position for anatomy j given the parameters of anatomy i (i.e., θ_i), we have:

$$\mathbf{p_{j|i}} = \mathbf{R_i}\left(\frac{\mathbf{p_j} - \mathbf{p_i}}{\sigma_\mathbf{j}}\right), \tag{6}$$

where $\mathbf{p_s} \in \mathbb{R}^\mathbf{3}$ and $\sigma_s \in \mathbb{R}$ are the center and scale of anatomy s, respectively. Given a training set $\{\theta_s(k)\}_{k=1,..,N}$, where k is an index to a training sample, we can formulate a least squares optimization for the scale invariant conditional position as:

$$\mu_{\mathbf{p(j|i)}} = \underset{\mathbf{p_{j|i}}}{\operatorname{argmin}} \mathbf{J}(\mathbf{p_{j|i}}), \tag{7}$$

where

$$J(\mathbf{p_{j|i}}) = \frac{\mathbf{1}}{\mathbf{2}} \sum_\mathbf{k} \left(\mathbf{p_j(k)} - \mathbf{p_i(k)} - \sigma_\mathbf{j}(\mathbf{k})\mathbf{R_i^T(k)p_{j|i}}\right)^\mathbf{2}, \tag{8}$$

leading to

$$\mu_{\mathbf{p(j|i)}} = \frac{\mathbf{1}}{\mathbf{N}} \sum_\mathbf{k} \mathbf{R_i(k)} \left(\frac{\mathbf{p_j(k)} - \mathbf{p_i(k)}}{\sigma_\mathbf{j}(\mathbf{k})}\right). \tag{9}$$

Assuming a Gaussian distribution for $\mathbf{p_{j|i}}$, the position covariance can be computed as:

$$\mathbf{\Sigma_{p(j|i)}} = \frac{\mathbf{1}}{\mathbf{N}} \sum_\mathbf{k} (\mathbf{p_{j|i}(k)} - \mu_{\mathbf{p(j|i)}})(\mathbf{p_{j|i}(k)} - \mu_{\mathbf{p(j|i)}})^\mathbf{T}. \tag{10}$$

The estimation of the scale of anatomy i given the scale of anatomy j is denoted as

$$\sigma_{j|i} = \frac{\sigma_j}{\sigma_i}. \tag{11}$$

[4]This detection sequence was found to provide the largest reduction in the search space, but we omit the details due to space limitations of the paper.

Again, considering a Gaussian distribution for $\sigma_{j|i}$, we have

$$\mu_{\sigma(j|i)} = \frac{1}{N}\sum_k \frac{\sigma_j(k)}{\sigma_i(k)}$$
$$\Sigma_{\sigma(j|i)} = \frac{1}{N}\sum_k (\sigma(j|i) - \mu_{\sigma(j|i)})^2. \quad (12)$$

Finally, the estimation of the orientation of anatomy i given the orientation of anatomy j is denoted as

$$\mathbf{q_{j|i}} = \mathbf{q_i} + \mathbf{d_q}(\mathbf{q_j}, \mathbf{q_i}), \quad (13)$$

where $d_q(.)$ is a function that computes difference between quaternions (see Sec. 2.4). Considering a Gaussian distribution for $\mathbf{q_{j|i}}$, we have

$$\mu_{\mathbf{q(j|i)}} = \frac{1}{N}\sum_k d_q(\mathbf{q_j(k)} - \mathbf{q_i(k)}) \quad (14)$$
$$\Sigma_{\mathbf{q(j|i)}} = \frac{1}{N}\sum_k (\mathbf{q(j|i)} - \mu_{\mathbf{q(j|i)}})(\mathbf{q(j|i)} - \mu_{\mathbf{q(j|i)}})^{\mathbf{T}}$$

Given the parameter estimations above, the computation of the semi-local context probabilities are as follows:

$$P(\theta_j|\{y_l = 1, \theta_l\}_{l=1,..,L}, V) = \quad (15)$$
$$g([\frac{1}{L}\sum_l R_l \frac{\mathbf{p_j} - \mathbf{p_l}}{\sigma_l}, \frac{1}{L}\sum_l \sigma_j/\sigma_l, \frac{1}{L}\sum_l d_q(\mathbf{q_j}, \mathbf{q_l})];$$
$$[\mu_{\mathbf{p(j|\{l\})}}\mu_{\sigma(j|\{l\})}\mu_{\mathbf{q(j|\{l\})}}], \Sigma)$$

with

$$\Sigma = \begin{bmatrix} \Sigma_{\mathbf{p(j|\{l\})}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Sigma_{\sigma(j|\{l\})} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \Sigma_{\mathbf{q(j|\{l\})}} \end{bmatrix}$$

and

$$g(\mathbf{x}; \mu, \mathbf{\Sigma}) = \frac{\mathbf{1}}{(\mathbf{2\pi})^{\mathbf{7/2}}|\mathbf{\Sigma}|^{\mathbf{1/2}}} \exp -\frac{\mathbf{1}}{\mathbf{2}}(\mathbf{x} - \mu)^{\mathbf{T}}\mathbf{\Sigma}^{\mathbf{-1}}(\mathbf{x} - \mu)),$$

where $\theta_j = [\mathbf{p_j}, \sigma_\mathbf{j}, \mathbf{q_j}]$ is the parameter for anatomy j, l is an index to the previous L detections, and $[\mu_{\mathbf{p(j|\{l\})}}\mu_{\sigma(j|\{l\})}\mu_{\mathbf{q(j|\{l\})}}]$ is computed by taking the sample average of the estimations, and similarly for Σ.

Notice that with the use of semi-local context, the complexity of the detection algorithm is unaltered, but in practice we can search only at places where the semi-local context probability is above a threshold. We note that empirically, we can disregard places in the parameter space that are further than 2 times the covariance of the estimated Gaussian. In general, this reduces the search space at each search parameter dimension from M to $M^{\frac{1}{2}}$ (recall that $M \approx 100$). As a result, in practice the complexity of the detection was reduced from $O\left(M^{7L}\right)$ in its original form (2) to $O\left(M^7 + L \times M^{\frac{7}{2}}\right)$. Consequently, the use of semi-local and global context information makes this approach linearly scalable in terms of brain anatomies.

2.3. Sequential Sampling to Model Probability Distributions

In this section we describe the process of modeling the posterior classifiers necessary to compute $P(y_s|\theta_s, V)$. Recall that $\theta_s \in \mathbb{R}^7$, which makes a brute force training and detection processes innefficient. We observe that sequential sampling [10] or marginal space learning [24] provides an efficient training and detection approaches for high-dimensional search parameter spaces. The main idea is to break the original parameter space Ω into subsets of increasing dimensionality $\Omega_1 \subset \Omega_2 \subset ... \subset \Omega$ and then train classifiers for each subset, where the samples for training the classifier in Ω_n are bootstrapped from Ω_{n-1}, and the classifier in Ω_1 is trained using all possible samples.

There is no clear methodology on how to divide this space, so we defined this sequence of spaces empirically, but omit the details due to space constraints of this paper. Specifically, we assumed the following sequence: $\Omega_1 = \mathbf{p} \in \mathbb{R}^\mathbf{3}$, $\Omega_2 = [\mathbf{p_s}, \sigma_\mathbf{s}] \in \mathbb{R}^\mathbf{4}$, and $\Omega_3 = \Omega = [\mathbf{p_s}, \sigma_\mathbf{s}, \mathbf{q_s}] \in \mathbb{R}^\mathbf{7}$. The actual search space for training and detection in Ω_n is defined to be $\dim(\Omega_n) - \dim(\Omega_{n-1})$, where $\dim(\Omega_n)$ denotes the dimensionality of the Ω_n space. In each subspace we train a discriminative classifier using the PBT algorithm [20] (i.e., forming PBT_n for each Ω_n) due to its ability of representing multi-modality distributions in binary classification problems. This process results in a training and detection complexity figures of $O\left(M^3\right)$, where M is the number of quantized parameter values per dimension. We would like to emphasize that this represents an *astounding reduction* in terms of complexity of the original algorithm in (2). Sequential sampling and the use of contextual information reduce the complexity from $O\left(M^{7L}\right)$ to $O\left(M^3 + L \times M^{\frac{3}{2}}\right)$. This reduction allows for the detection of additional anatomies with little impact on the overal detection complexity.

2.4. Orientation Using Quaternions

The space of possible orientations is usually represented with the three Euler angles [24]. Euler angles are easy to implement and understand, but they have several drawbacks to represent the orientation space. First, a uniform step size over possible Euler angles *does not* generate a uniform sampling in the space of orientations, which makes Euler angles impractical for the uniform sampling of the space of orientations. Second, the representation for each orientation is not unique, which makes difficult to define a similarity measure between two orientations expressed in Euler angles. In other words, Euler angles are a chart of the space of orientations with singularities (i.e., non-smooth). Consequently, two similar orientations might have very different Euler angles, which makes it difficult to compute statistics and distances. For example, if we select the ZXZ-convention with (γ, β, α) as the three Euler angles, we have a singularity along the line $\beta = 0$. The triplet $(0.7, 0.0, 0.3)$ gives the same orientation as the triplet $(0.0, 0.0, 1.0)$.

We use the concepts on quaternions proposed by Karney et al. [8] for molecular modeling to represent the space of orientations in 3D. All the problems exposed above are solved using unitary quaternions to express orientations. Each orientation can be defined as a point in the hypersphere $\mathbb{S}^3 = \{p \in \mathbb{R}^4 | \|p\|_2 = 1\}$ with opposite points identified[5]. This equivalence relation defines the space of orientations as the following quotient space:

$$SO(3) = \{\mathbb{S}^3/\{q, -q\} | q = [q_1, q_2, q_3, q_4] \in \mathbb{R}^4; \parallel q \parallel_2^2 = 1\},$$

where the operator / denotes the quotient space given by the identification of $\{q \sim -q\}$ in $\mathbb{S}^3$. We use mainly two properties from quaternions. First, composition of two rotations can be computed as a multiplication of two quaternions. If $R(q)$ with $q \in SO(3)$ represents one orientation and $R(p)$ with $p \in SO(3)$ represents another, then $R(p) \circ R(q) = p \cdot \overline{q}$ where $\overline{q}$ is the conjugate of q[6]. Second, there is a distance preserving map between $SO(3)$ and a ball in $\mathbb{R}^3$. This map allows us to use in $SO(3)$ standard statistical tools from $\mathbb{R}^3$.

Each quaternion can also be expressed as $q = [cos(\theta/2)\, v \cdot sin(\theta/2)] \in SO(3)$ with $v \in \mathbb{R}^3$ s.t. $\parallel v \parallel_2 = 1$ and $\theta \in (-\pi, \pi)$. The intuition behind this is that v represents the axis of rotation and θ the angle of rotation around that axis. Then, the definition of the distance preserving map is the following:

$$\begin{aligned} f : SO(3) &\longrightarrow \mathbb{R}^3 \\ q &\longmapsto u \text{ with } u \parallel v \text{ and } \|u\| = \left(\frac{|\theta| - \sin(|\theta|)}{\Pi}\right)^{\frac{1}{3}} \end{aligned} \tag{16}$$

The same way we can 'flatten' a hemisphere into a disc in $\mathbb{R}^2$ preserving geodesic distances, Eq. 16 'flattens' the quotient space into a ball in $\mathbb{R}^3$. All the details can be found in [8].

Using the two properties explained above, it is easy to manipulate orientations and to compute statistics and metrics in the space of orientations. We have that $d_q(\mathbf{q_j}, \mathbf{q_i})$ in Eq. 13 can be defined as

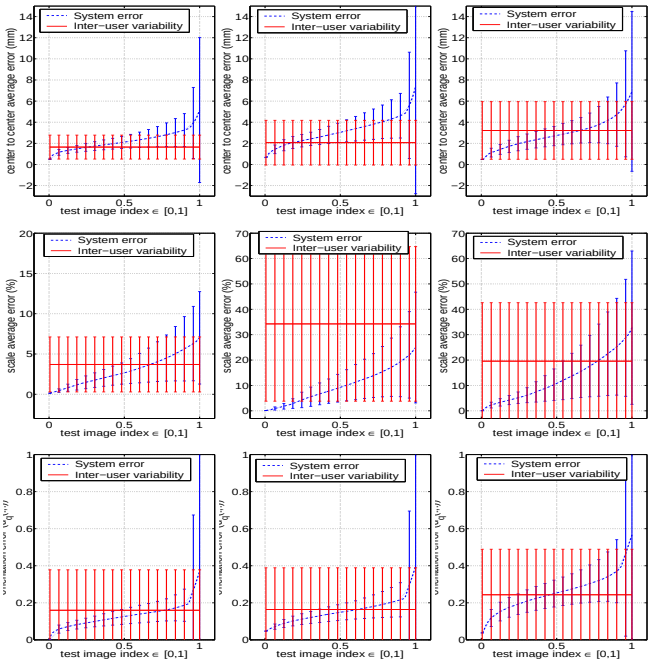
$$d_q(\mathbf{q_j}, \mathbf{q_i}) = \|\mathbf{f}(\mathbf{q_j}) - \mathbf{f}(\mathbf{q_i})\|_\mathbf{2}, \tag{17}$$

where f is defined in Eq. 16.

The authors in [8] also provide a method to sample the space of orientations uniformly with different angle precision. The explanation of such a method is non-trivial and it is out of the scope of this paper. We store in memory the sampling points for different resolutions since they are not easy to calculate on-the-fly. For example, to achieve 10° accuracy only 7416 samples are needed using the method in [8]. Using constant step size in Euler angles $36*36*18 = 23328$ samples are needed. Since we need to sample the complete space of orientations in 3DUS, quaternions bring enormous savings to this task.

[5] If $p \in \mathbb{S}^3$ then the opposite point of p is $-p$, which also belongs to $\mathbb{S}^3$.

[6] $\overline{q} = [q_1, -q_2, -q_3, -q_4]$



(a) cerebellum (b) cist. magna (c) lat. ventricles

Figure 3. Comparison between system error and inter-user variability. The horizontal axis displays an index (from 0 to 1) of the test set volumes (sorted in ascending order in terms of the system error), and the vertical axes show the error for position (row 1), scale (row 2), and orientation (row 3) computed according to (18). The solid curve shows the average inter-user variability, while the dotted curve displays the average error of the system up to the test image indicated by the index in the horizontal axis (error bars shows the standard deviation (18)).

3. Training Protocol

Details of the training protocol for the algorithm explained in Sec. 2 are provided in this section. We collected 240 volumes with expert annotations of cerebellum, cisterna magna, and lateral ventricles (see Fig. 2). Volumes have an average size of $250 \times 200 \times 150$. We also built the annotation for the center of the brain using the same annotation plane as the Cerebellum, and drawing a line through the midline of the brain (see Fig. 2). The training volumes for each anatomy are obtained by building a sub-volume around the annotation of size k times bigger the annotation length (note that $k = 2$ for CER, $k = 7$ for CM, $k = 5$ LV, and $k = 1.5$ for CB) – see Fig. 2. The PBT_1, or marginal classifier for position, is trained with positive samples formed by a box around the center location of the annotated anatomy with fixed size and oriented according to the volume orientation (i.e., not according to the annotation orientation). The negative samples are formed with boxes from positions δ_p away from this center (here $\delta_p = 3$ voxels). The features used for PBT_1 were the 3-D Haar features [24, 22] because of the high efficiency in its computation using integral volumes. The classifier for position and scale, PBT_2, is trained with positive samples formed by a box around the center of

the anatomy with size proportional to the length of annotation, but oriented according to the volume orientation. The negative samples are boxes δ_p away from the center and δ_s away in terms of scale (we consider $\delta_s = 2$). Finally, for PBT_3, or orientation classifier, we build the positive training samples with boxes located at the anatomy center, proportional to scale and at the correct orientation. Negative samples are boxes δ_p, δ_s, and δ_q away, where $\delta_q = 0.2$. For $PBT_{2,3}$ we used the steerable features [24] because of the efficiency of their computation. The main advantage is that, differently of the 3-D Haar features, it is not necessary to perform volume rotations to compute these features. Recall that the training of PBT_n uses bootstrapped samples from PBT_{n-1}. The process explained above produces the discriminative classifiers $P(y_s = 1|\theta_s, V)$. The semi-local context parameters are learned generatively as detailed in Sec. 2.2.

4. Results

In this section, we show the results of an experiment using 200 volumes for training and 40 volumes for testing, where the training and test volumes were randomly selected and there is no overlap between the training and test sets. We compared the results produced by the system with the results from an inter-user variability experiment conducted with two OBGYN experts who measured the Cerebellum, Cisterna Magna, and Lateral Ventricles on the same volumes (see Fig. 3). The average and standard deviation of the inter-user variability and system error for position, scale, and orientation are respectively computed as follows:

$$\begin{array}{ll} \mu_{\mathbf{p}}=\frac{1}{N}\sum_{i=1}^{N}\|\mathbf{p1_i}-\mathbf{p2_i}\|, & \sigma^2_{\mathbf{p}}=\frac{1}{N}\sum_{i=1}^{N}(\|\mathbf{p1_i}-\mathbf{p2_i}\|-\mu_{\mathbf{p}})^2, \\ \mu_{\sigma}=\frac{1}{N}\sum_{i=1}^{N}|\sigma 1_i-\sigma 2_i|, & \sigma^2_{\sigma}=\frac{1}{N}\sum_{i=1}^{N}(|\sigma 1_i-\sigma 2_i|-\mu_{\sigma})^2, \\ \mu_{\mathbf{q}}=\frac{1}{N}\sum_{i=1}^{N}|d_q(\mathbf{q1_i},\mathbf{q2_i})|, & \sigma^2_{\mathbf{q}}=\frac{1}{N}\sum_{i=1}^{N}(|d_q(\mathbf{q1_i},\mathbf{q2_i})|-\mu_{\mathbf{q}})^2, \end{array} \quad (18)$$

where N is the number of volumes for testing and $d_q(.,.)$ is defined in (17). We assumed that one of the user experts produced the ground truth results, so in (18), the index 1 denotes ground truth (i.e., one of the users) while 2 indicates either the measurements by the other user for the case of the inter-user variability or the system automatic measurements for the computation of the system error. Notice in Fig. 3 that the average error of the automatic results produced by the system is within the range of inter-user variability for all cases except for 10% to 20% of the cases for Cerebellum and Cisterna Magna position. Empirically, we obtained the best trade-off between robustness (to imaging variations, noise, and pose variance) and accuracy by running the system on a pyramid of volumes where the coarse scale is 4mm/voxel (isotropic) and the finest scale is 2mm/voxel. Consequently, the error results produced by the system have a different scale than the inter-user variability that partially explains this discrepancy.

Fig. 4 shows several results of the system and a comparison with the user measurements (this user was the one assumed to produced the ground truth measurements).

5. Conclusion and Future Work

We presented an approach that is capable of automatically indexing 3-D ultrasound volumes of fetal heads using semantic keywords, which represent fetal anatomies. The automatic index involves the display of the correct standard plane for visualizing the requested anatomy and the biometric measurement according to the guidelines of the International Society of Ultrasound in Obstetrics and Gynecology [12]. Our approach represents the first method that is able to retrieve anatomies in ultrasound volumes based on semantic keywords. We show this system working with three brain anatomies, but we are currently expanding it to work with tens of brain anatomies, and ultimately our goal is to index all important fetal body anatomies in 3DUS. In order to achieve this goal we propose a novel principled probabilistic model that combines the use of discriminative/generative classifiers with global and semi-local context. Results in a large experimental set-up show that our system produces biometric measurements and show standard planes that are within the inter-user variability. Finally, this system currently runs under 10 seconds on a standard dual core computer running at 1.7GHz, but we anticipate that with standard code optimization, this system will run under 3 seconds.

Acknowledgements: We would like to thank Dr. Matthias Scheier for providing volume annotations.

References

[1] B. Benacerraf, T. Shipp, and B. Bromley. How sonographic tomography will change the face of obstetric sonography: A pilot study. *J Ultrasound Med*, 24(3):371–378, 2005. 1

[2] G. Carneiro and D. Lowe. Sparse flexible models of local features. *ECCV*, 2006. 4

[3] F. F. Correa, C. Lara, J. Bellver, J. Remoh, A. Pellicer, and V. Serra. Examination of the fetal brain by transabdominal three-dimensional ultrasound: potential for routine neurosonographic studies. *Ultrasound in Obstetrics and Gynecology*, 27(5):503–508, 2006. 1

[4] B. Benacerraf et al. Three- and 4-dimensional ultrasound in obstetrics and gynecology - proceedings of the american institute of ultrasound in medicine consensus conference. *J Ultrasound Med*, 24:1587–1597, 2005. 1

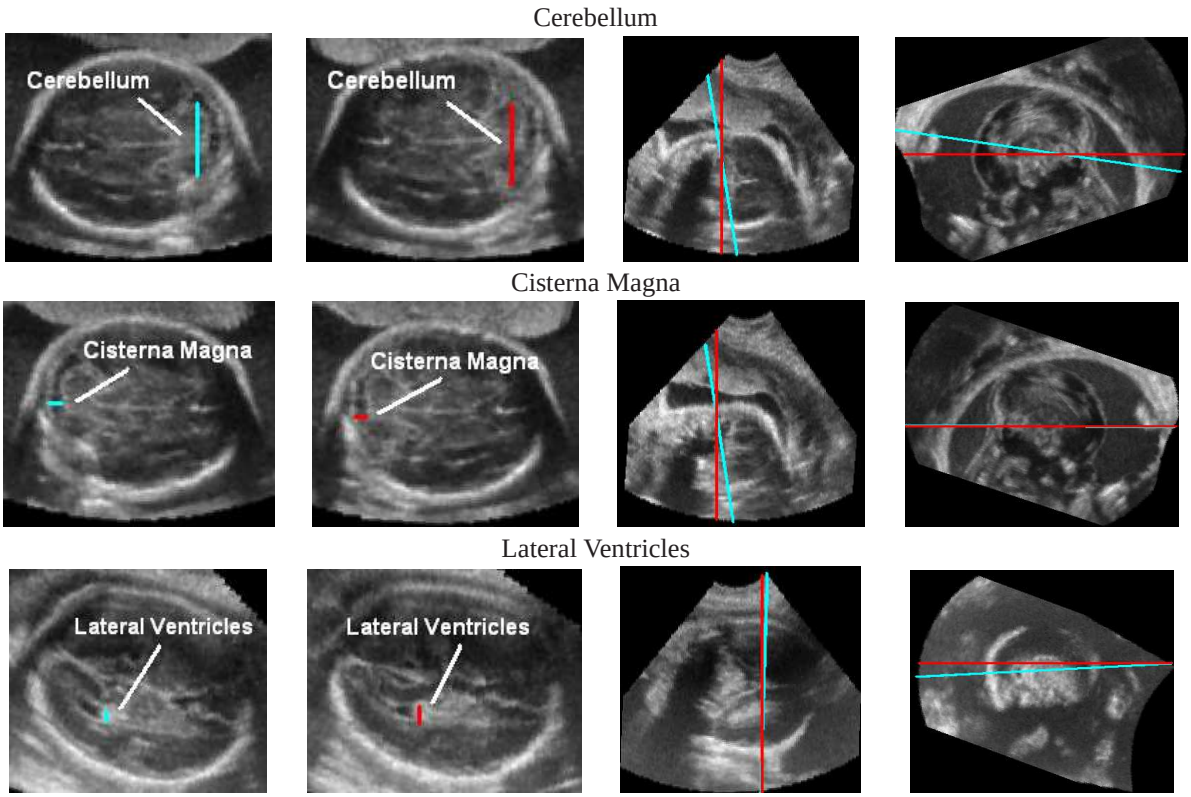
[5] M. Gooding, S. Kennedy, and J. A. Noble. Volume reconstruction from sparse 3d ultrasonography. *MICCAI*, 2003. 2

[6] W. Hong, B. Georgescu, X. Zhou, S. Krishnan, Y. Ma, and D. Comaniciu. Database-guided simultaneous multi-slice 3d segmentation for volumetric data. *ECCV*, 2006. 2

[7] A. Johnson and M. Hebert. Using spin images for efficient object recognition in cluttered 3d scenes. *IEEE TPAMI*, 21(5), 1999. 2

[8] C. F. F. Karney. Quaternions in molecular modeling. *J.MOL.GRAPH.MOD.*, 25:595–604, 2006. 6

[9] G. Michailidis, P. Papageorgiou, and D. Economides. Assessment of fetal anatomy in the first trimester using two- and three-dimensional ultrasound. *British Journal of Radiology*, 75:215–219, 2002. 1



a) Detection b) Annotation c) Coronal plane d) Sagittal plane

Figure 4. Detection (blue) and ground truth (red) annotations shown on the transverse plane (columns 1 and 2, respectively). Detection (blue) and ground truth (red) cross section planes on the sagittal and coronal planes (columns 3 and 4, respectively). Only one dataset per anatomy is shown due to space limitation. We have similar images for each volume.

[10] P. Moral, A. Doucet, and G. Peters. Sequential monte carlo samplers. *J. R. Statist. Soc. B*, 68:411436, 2006. 2, 5

[11] M. Necas. Musculoskeletal symptomatology and repetitive strain injuries in diagnostic medical sonographers: A pilot study in washington and oregon. *Journal of Diagnostic Medical Sonography*, 12:266–273, 1996. 1

[12] T. I. S. of Ultrasound in Obstetrics and Gynecology. Sonographic examination of the fetal central nervous system: guidelines for performing the basic examination and the fetal neurosonogram. *Ultrasound in Obstetrics and Gynecology*, 29:109–116, 2007. 1, 2, 7

[13] M. Oren, C. Papageorgiou, P. Sinha, E. Osuna, and T. Poggio. Pedestrian detection using wavelet templates. *IEEE CVPR*, 1997. 2

[14] I. Pike, A. Russo, J. Berkowitz, J. Baker, and V. A. Lessoway. The prevalence of musculoskeletal disorders among diagnostic medical sonographers. *Journal of Diagnostic Medical Sonography*, 13:219–227, 1997. 1

[15] A. Rabinovich, A. Vedaldi, C. Galleguillos, E. Wiewiora, and S. Belongie. Objects in context. *ICCV*, 2007. 3

[16] R. Rohling, A. Gee, and L. Berman. Automatic registration of 3-d ultrasound images. *ICCV*, pages 298–303, 1998. 2

[17] B. H. Romeny, B. Titulaer, S. Kalitzin, G. Scheffer, F. Broekmans, J. Staal, and E. te Velde. Computer assisted human follicle analysis for fertility prospects with 3d ultrasound. *Proceedings of the International Conference on Information Processing in Medical Imaging*, pages 56–69, 1999. 2

[18] R. Shekhar and V. Zagrodsky. Mutual information-based rigid and nonrigid registration of ultrasound volumes. *IEEE Transactions on Medical Imaging*, 21(1):9–22, 2002. 2

[19] A. Torralba. Contextual priming for object detection. *IJCV*, 53(2):169–191, 2003. 2, 3

[20] Z. Tu. Probabilistic boosting-tree: Learning discriminative methods for classification, recognition and clustering. *ICCV*, pages 1589–1596, 2005. 2, 4, 5

[21] Z. Tu, K. Narr, P. Dollar, I. Dinov, P. Thompson, and A. Toga. Brain anatomical structure segmentation by hybrid discriminative/generative models. *Transactions on Medical Imaging*, 2007. 2

[22] Z. Tu, X. Zhou, D. Comaniciu, and L. Bogoni. A learning based approach for 3d segmentation and colon detagging. *ECCV*, 2006. 2, 6

[23] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. *CVPR*, pages 511–518, 2001. 2

[24] Y. Zheng, A. Barbu, B. Georgescu, M. Scheuering, and D. Comaniciu. Fast automatic heart chamber segmentation from 3d ct data using marginal space learning and steerable features. *ICCV*, 2007. 2, 5, 6, 7