Automatic Fetal Measurements in Ultrasound Using Constrained Probabilistic Boosting Tree

Gustavo Carneiro[1], Bogdan Georgescu[1], Sara Good[2], and Dorin Comaniciu[1]

[1] Siemens Corporate Research, Integrated Data Systems Dept., Princeton,NJ,USA,
[2] Siemens Medical Solutions, Innovations Ultrasound Div., Mountain View,CA,USA.

Abstract. Automatic delineation and robust measurement of fetal anatomical structures in 2D ultrasound images is a challenging task due to the complexity of the object appearance, noise, shadows, and quantity of information to be processed. Previous solutions rely on explicit encoding of prior knowledge and formulate the problem as a perceptual grouping task solved through clustering or variational approaches. These methods are known to be limited by the validity of the underlying assumptions and cannot capture complex structure appearances. We propose a novel system for fast automatic obstetric measurements by directly exploiting a large database of expert annotated fetal anatomical structures in ultrasound images. Our method learns to distinguish between the appearance of the object of interest and background by training a discriminative constrained probabilistic boosting tree classifier. This system is able to handle previously *unsolved* problems in this domain, such as the effective segmentation of fetal abdomens. We show results on fully automatic measurement of head circumference, biparietal diameter, abdominal circumference and femur length. Unparalleled extensive experiments show that our system is, on average, close to the accuracy of experts in terms of segmentation and obstetric measurements. Finally, this system runs under half second on a standard dual-core PC computer.

1 Introduction

Accurate fetal ultrasound measurements are one of the most important factors for high quality obstetrics health care. Common fetal ultrasound measurements include: bi-parietal diameter (BPD), head circumference (HC), abdominal circumference (AC), and femur length (FL). These measures are used to estimate the gestational age (GA) of the fetus [11] and are an important diagnostic tool. Although prevalent in clinical setting, the manual measurement by specialists of BPD, HC, AC, and FL present the following issues: 1) the quality of the measurements are user-dependent, 2) the exam can take more than 30 minutes, and 3) specialists can suffer from Repetitive Stress Injury (RSI) due to these lengthy exams. The automation of these ultrasound measures has the potential of improving productivity and patient throughput, enhancing accuracy and consistency of measurements, and reducing the risk of RSI to specialists.

In this paper we present an *on-line* system that targets the accurate and robust detection *and* segmentation of fetal head, abdomen, and femur in ultrasound images. The segmentation information is then used to compute BPD,

HC, AC, and FL. Our approach directly exploits the expert annotation of fetal anatomical structures in large databases of ultrasound images [3] to train a novel discriminative appearance classifier of simple image features derived from the probabilistic boosting tree classifier (PBT) [12]. Our method can handle two previously unsolved issues in the domain of fetal ultrasound imaging. First, our system is able to provide an accurate segmentation of the fetal abdomen. Second, the approach was designed to be totally automatic, so the user does not need to provide any initial guess. The only inputs to the system is the image and the measurement to be performed (BPD, HC, AC, or FL). Extensive experiments show that, on average, there is practically no difference between the measurement produced by our system and the annotation made by experts for the four fetal measurements mentioned above. Moreover, the algorithm runs under half second on a standard dual core PC computer.

2   Literature Review

The detection and segmentation of fetal anatomical structures in ultrasound images have been studied by several researchers [1, 4, 5, 9]. An important common point among these papers is that they heavily exploit the low-level structures of the imaging of the fetal anatomy, such as edge or texture, which is usually effective at segmenting heads and femurs due to strong signal responses produced by the imaging of the skull and femur. However, the segmentation of abdomens (see Fig. 3) represents a more challenging problem, where such set of low-level structures has proved to be extremely hard to find. The most promising approaches by Chalana et al. [1, 9] and by Jardim and Figueiredo [5] share the idea of describing the segmentation as an optimization process. They also share the issue that the optimization can get stuck at local minima, which may produce low quality segmentation results in images presenting noise and/or missing data. A constraint presented by both papers is the need of an initial guess by the user, yielding a semi-automatic system. Nevertheless, these papers present results that are of comparable accuracy as sonographers in small datasets (less than 50 images) of fetal heads and femurs. A common point among the papers above is that they did not work on the segmentation of fetal abdomen.

In echocardiography, the detection and segmentation of the left ventricle of the heart in ultrasound images have produced similar algorithms based on an optimization process [14]. The issue of the arbitrary initial condition has been handled with the use of level sets [8], and the robustness to noise and missing data has been managed by the use of a shape influence term [10]. However, these methods still present some basic issues, such as the under utilization of the appearance model because of the use of relatively simple parametric models for the appearance term in the optimization function. Also, these techniques tend to work well whenever image gradients separate the sought anatomical structure, but that might not be the case for complex anatomies (e.g., fetal abdomens).

The method we propose in this paper is aligned with the state-of-the-art detection and high-level segmentation methods proposed in computer vision and machine learning. These methods involve user annotated training data, discriminant classifiers of local regions, and a way to combine the results of those local classifiers [6, 3]. We exploit the database-guided segmentation paradigm [3] in the domain of fetal ultrasound images. This domain presents common issues en-
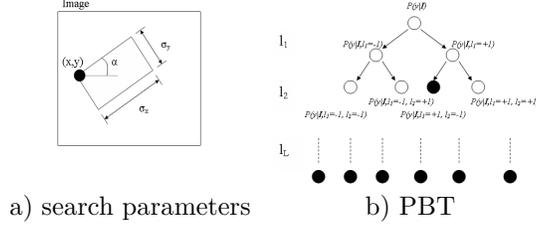
Fig. 1. Image region with five parameters (a) and the PBT tree structure (b).

countered in ultrasound images, such as large amount of noise, signal drop-out and large variations between the appearance, configuration and shape of the anatomical structure. However, our method has to handle new challenges presented by fetal ultrasound images, such as the extreme appearance variability of the fetal abdomen imaging as well as to be easily generalized to all anatomical structures. In order to cope with these new challenges, we constrain the recently proposed probabilistic boosting tree classifier [12] to limit the number of nodes present in the binary tree, and also to divide the original classification problem into stages of increasing complexity.

3 Automatic Measurement of Fetal Anatomy

In this section we define the segmentation problem and explain the training and detection processes used for the process of automatic segmentation and measurement of fetal anatomies.

3.1 Problem Definition

The ultimate goal of our system is to compute the probability that an image region contains the sought structure of interest. An image region is represented by a set of N image features (here we use the Haar wavelet features [7, 13] computed using integral images), so we define the vector $I \in \Re^N$ as follows:

$$I = f(\theta), \tag{1}$$

where $\theta = [x, y, \alpha, \sigma_x, \sigma_y]$ with the parameters (x, y) representing the top left region position in the image, α denoting orientation, (σ_x, σ_y), the region scale (see Fig. 1-(a)), and the function $f(\theta)$ computes the N image features in the region. Note that with these parameters, the ellipsoidal measurements for HC and AC are computed with a set of M points as follows: $\{(x + r\sigma_x \cos\alpha\cos\gamma + r\sigma_y \sin\alpha\sin\gamma, y - r\sigma_x \sin\alpha\cos\gamma + r\sigma_y \cos\alpha\sin\gamma)\}$, where $\gamma = j\frac{2\pi}{M}$ with $j = \{0, ..., M-1\}$, $r = 0.75$, and the line measurements for BPD and FL are defined by the two end points of line:$\{(x + r\sigma_i \cos(\alpha + \gamma), y + r\sigma_i \cos(\alpha + \gamma))\}$, where $r = \{-0.75, +0.75\}$, for FL, $i = x$ and $\gamma = 0$, and for BPD, $i = y$ and $\gamma = \pi/2$. A classifier then defines the following function: $P(y|I)$, where $y \in \{-1, +1\}$ with $P(y = +1|I)$ representing the probability that the image region I contains the structure of interest (i.e., a positive sample), and $P(y = -1|I)$, the probability that the image region I contains background information (i.e., a negative sample). Notice that the main goal of the system is to determine $\theta^* = \arg\max_\theta P(y|f(\theta))$. Therefore, our task is to train a discriminative classifier that minimizes the probability of mis-classification.

3.2 Constrained Probabilistic Boosting Tree

The classifier used for the anatomical structure detection is derived from the probabilistic boosting tree classifier (PBT) [12]. Training the PBT involves the recursive construction of a tree, where each of its nodes represents a strong classifier. The input training set for each node is divided into two sets (left or right) according to the result provided by the learned strong classifier. Each new set is then used to train the left and right sub-trees recursively. The posterior probability that a sample is positive is computed as follows [12]:

$$P(y = +1|\mathbf{I}) = \sum_{l_1,l_2,...,l_n} P(y = +1|l_n,...,l_1,\mathbf{I})...P(l_2|l_1,\mathbf{I})P(l_1|\mathbf{I}), \qquad (2)$$

where n is the total number of nodes of the tree (see Fig. 1-(b)). The original PBT classifier presents a problem: if the classification is too hard, two problems may occur: a) overfit of the training data in the nodes close to the leaves and b) long training and detection procedures. As a result, in order to reduce the training and detection times and improve the generalization ability of the classifier, we propose the Constrained PBT (CPBT) algorithm, which constrains the original PBT training process as follows: a) instead of having only one classifier, we train a sequence of classifiers with increasing complexity; and b) given that the classification problem of each classifier is simpler than the original problem, the height and the number of tree nodes are constrained.
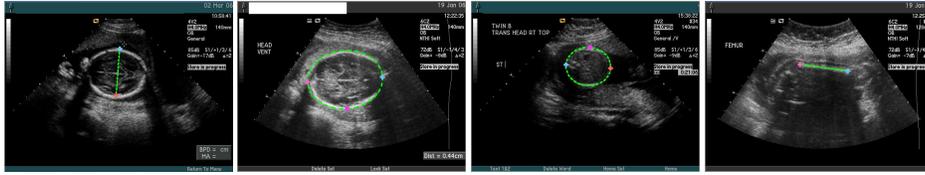

Fig. 2. Expert annotation of (left to right): BPD, HC, AC, and FL.

Algorithm 1: Training algorithm.

Data : Set of training images with respective true positive parameter $\{(I, \theta_I^+)_i\}_{i=1,..,M}$
Coarse search sampling interval δ_{coarse}
Height H_{coarse} of the Coarse classifier tree and total number of nodes N_{coarse}
Fine search sampling interval δ_{fine}
Height H_{fine} of the Fine classifier tree and total number of nodes N_{fine}

$\mathcal{I}^+ = \emptyset$ and $\mathcal{I}^- = \emptyset$
for $i = 1, ..., M$ **do**
    Generate P positives samples by randomly sample the parameter space with
    $\tilde{\theta}_I^+ \in [\theta_I^+ - \delta_{\text{coarse}}, \theta_I^+ + \delta_{\text{coarse}}]$ and add these samples to the set of positive samples $\mathcal{I}^+$
    Generate N negative samples by randomly sample the parameter space with
    $\tilde{\theta}_I^- \notin [\theta_I^+ - \delta_{\text{coarse}}, \theta_I^+ + \delta_{\text{coarse}}]$ and add these samples to the set of negative samples $\mathcal{I}^-$
end

Train coarse classifier of height H_{coarse} and number of nodes N_{coarse} using $\mathcal{I}^+$ and $\mathcal{I}^-$.
$\mathcal{I}^+ = \emptyset$ and $\mathcal{I}^- = \emptyset$
for $i = 1, ..., M$ **do**
    Generate P positives samples by randomly sample the parameter space with
    $\tilde{\theta}_I^+ \in [\theta_I^+ - \delta_{\text{fine}}, \theta_I^+ + \delta_{\text{fine}}]$ and add these samples to the set of positive samples $\mathcal{I}^+$
    Generate N negative samples by randomly sample the parameter space with
    $\tilde{\theta}_I^- \notin [\theta_I^+ - \delta_{\text{fine}}, \theta_I^+ + \delta_{\text{fine}}]$ and $\tilde{\theta}_I^- \in [\theta_I^+ - \delta_{\text{coarse}}, \theta_I^+ + \delta_{\text{coarse}}]$ and add these samples to the set of negative samples $\mathcal{I}^-$
end

Train fine classifier of height H_{fine} and number of nodes N_{fine} using $\mathcal{I}^+$ and $\mathcal{I}^-$.
Result : Coarse and fine classifiers.

3.3 Training a Constrained PBT

Figure 2 shows expert annotations of fetal measurements. Note that this annotation explicitly defines the parameter θ_I^+ for the true positive sample of the training image (1). Figure 3 shows image regions returned by sampling the parameter space with θ_I^+ for several training images.

As mentioned in Sec. 3.2, the training processes involves a two stage classification problem of increasing complexity (see Alg. 1). The first stage, referred to as coarse stage, is robust to false negatives, but it accepts a relatively large number of false positives, while the second stage, which is called fine stage, is more selective, being robust to false positives.

Fig. 3. Examples of the training set for BPD and HC (first three images), AC (images 4-6), and FL (last three images).

3.4 Detection

The detection algorithm also runs in two stages, as described in Algorithm 2. The coarse detection samples the search space uniformly using the δ_{coarse} as the sampling interval, while the fine detection searches the hypotheses selected from the coarse search at smaller intervals of δ_{fine}.

Algorithm 2: Detection algorithm.

Data : Test image and anatomy to be detected
Coarse and fine classifiers
Measuremet to be performed (BPD, HC, AC, or FL)
$\mathcal{H}_{\text{coarse}} = \emptyset$
for $\theta = [0, 0, 0, 0, 0] : \delta_{coarse} : \max(\textit{search space})$ **do**
    Sample image region $I = f(\theta)$ (1)
    Compute $P(y = +1|I)$ (2) using coarse classifier
    $\mathcal{H}_{\text{coarse}} = \mathcal{H}_{\text{coarse}} \bigcup (\theta, P(y = +1|I))$
end
Assigned the top H hypotheses from $\mathcal{H}_{\text{coarse}}$ in terms of $P(y = +1|I)$ to $\mathcal{H}_{\text{fine}}$
for $i = 1, ..., H$ **do**
    Assume $(\theta_i, P_i) = i^{th}$ element of $\mathcal{H}_{\text{fine}}$
    for $\theta = (\theta_i - \delta_{coarse}) : \delta_{fine} : (\theta_i + \delta_{coarse})$ **do**
        Sample image region $I = f(\theta)$ (1)
        Compute $P(y = +1|I)$ (2) using fine classifier
        $\mathcal{H}_{\text{fine}} = \mathcal{H}_{\text{fine}} \bigcup (\theta, P(y = +1|I))$
    end
end
Select the top hypothesis from $\mathcal{H}_{\text{fine}}$ in terms of $P(y = +1|I)$.
Result : Parameter θ of the top hypothesis.

3.5 Training Results

We have $1,426$ expert annotated training samples for head, $1,168$ for femur, and $1,293$ for abdomen. A coarse and a fine CPBT classifiers were trained, each with six levels. We are interested in determining the tree structure of the classifier,

where we want to constrain the tree to have the fewest possible number of nodes without affecting the classifier performance. Recall from Sections 3.3 and 3.4 that the fewer the number of nodes, the more efficient the training and the detection algorithms. Here, we compare the performance of the full binary tree against a tree constrained to have only one child per node. The number of weak classifiers in each node is arbitrarily set at between 10 and 300 depending on the tree structure (full binary tree classifiers usually uses fewer weak classifiers to avoid over-fitting). The sampling interval for the training algorithm 1 is set as $\delta_{\text{coarse}} = [20, 20, 20, 20, 20]$, and $\delta_{\text{fine}} = [10, 10, 10, 10, 10]$. These two vectors contain the variation for each dimension of the parameter space (i.e., position (x, y) in pixels, orientation α in angles, and scale (σ_x, σ_y) in pixels). Note that these sampling intervals were set based on experiments ommited in this paper due to lack of space. Finally in Algorithm 1, the number of additional positives per image $P = 100$ and the number of negatives per image $N = 1000$.

In Fig. 4 we see the measurement errors for HC and BPD in the training for the constrained tree and the full binary tree. Assuming that the GT contains the expert annotation and DT denotes the automatic measurement produced by the system, the error is computed as:

$$\text{error} = |GT - DT|/GT. \tag{3}$$

Notice that the performance of the constrained tree is better than that of the full binary tree for the parameters. This is explained by the fact that the constrained tree is more regularized and should be able to generalize better than the full binary tree. Another key advantage of the constrained tree is the efficiency in training. For the cases above, the training process for the full binary tree takes between seven to ten days, while for the constrained tree the whole training takes two to four days on a standard PC computer. Hence, a constrained tree classifier should be used in the experiments.
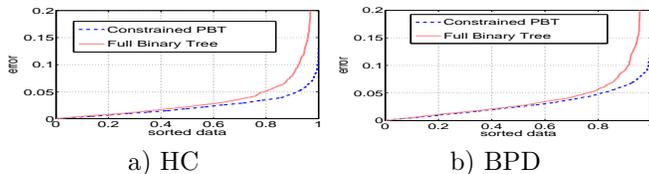

a) HC   b) BPD

Fig. 4. Training error (3) comparison between cascade and full binary tree.

4 Experimental Results

In this section we show qualitative and quantitative results of the database-guided image segmentation based on the CPBT classifier. Using the detection algorithm described in Sec. 3.5, which was trained using the training set listed in Sec. 3.5 for each anatomy, we compute the error in a test set composed of 177 ultrasound images of fetal head, 183 of fetal abdomen, and 171 of fetal femur. Fig. 5 shows the error in ascending order for each anatomy in the training and test sets. Table 1 displays the median and mean errors for the training
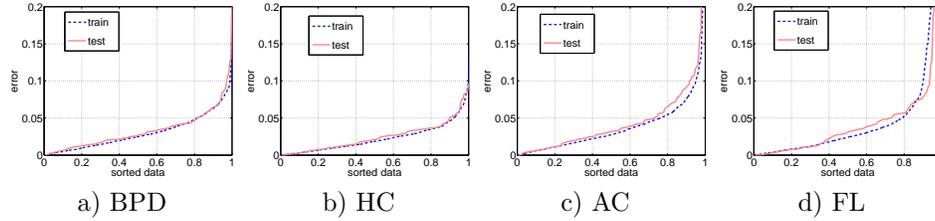
| a) BPD | b) HC | c) AC | d) FL |

Fig. 5. Error (3) in ascending order for each anatomy in the training and test sets. The horizontal axis is normalized to vary from zero to one.

and test sets. The running time for this algorithm is under half second for all measurements on a standard dual-core PC computer. Finally, Fig. 6 displays a few segmentation results produced by our method.

Table 1. Median and mean errors (3) of each one of the measurements.

Measurement	BPD	HC	AC	FL
median error in training set	0.0249	0.0187	0.0277	0.0241
median error in test set	0.0269	0.0211	0.0319	0.0316
mean error in training set	0.0306	0.0225	0.0399	0.0668
mean error in test set	0.0333	0.0247	0.0479	0.0588

This system was extensively tested in a clinical setting of ultrasound examinations. In this experiment, we noticed that 20% of the cases show a relatively large error. Since these cases can be considered to be outliers, we left them out of the following results. The results showed that in the remaining 80% of the test cases, the system produced an average error of 0.0265 (3) with respect to a ground truth measurement computed as the average of the measurement of 15 experts in hundreds of images presenting fetal heads, femurs, and abdomens. It is important to mention that the average error between the users' measurements and the same ground truth was exactly 0.0265.

5   Conclusions

We presented a system that automatically measures the BPD and HC from ultrasound images of fetal head, AC from images of fetal abdomen, and FL in images of fetal femur. Our system uses a large database of expert annotated images in order to train a Constrained Probabilistic Boosting Tree classifier. We showed that this system is capable of handling previously unsolved issues in the domain of fetal ultrasound imaging, such as the effective abdomen segmentation, and completely automated measurement procedure without user assistance. The results show that our system produces accurate results, and the clinical evaluation shows results that are, on average, close to the accuracy of sonographers. Moreover, the algorithm is extremely efficient and runs in under half second on a standard dual-core PC computer. Finally, the clinical evaluation also showed a seamless integration of our system into the clinical workflow.
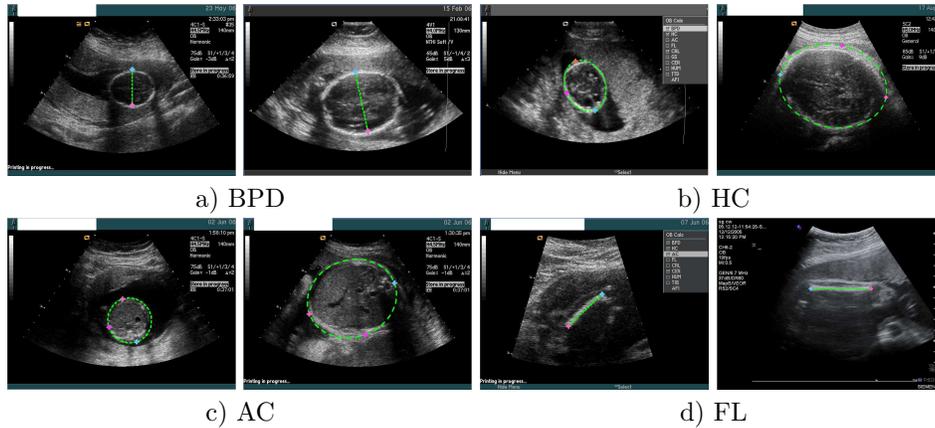
a) BPD   b) HC

c) AC   d) FL

Fig. 6. Detection and segmentation results.

References

1. V. Chalana, T. Winter II, D. Cyr, D. Haynor, Y. Kim. Automatic fetal head measurements from sonographic images. In *Acad Radiology*, 3 (8), pp. 628-635, 1996.
2. J. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: a statistical view of boosting. *Annals of Statistics*, 28 (2), pp. 337-374, 2000.
3. B. Georgescu, X. Zhou, D. Comaniciu, and A. Gupta. Database-guided segmentation of anatomical structures with complex appearance. In *IEEE CVPR*, 2005.
4. C. Hanna and A. Youssef. Automated measurements in obstetric ultrasound images. In *ICIP*, 1997.
5. S. Jardim and M. Figueiredo. Segmentation of fetal ultrasound images *Ultrasound in Medicince and Biology*, 31 (2), pp. 243-250, 2005.
6. A. Levin, Y. Weiss, Learning to combine bottom-up and top-down segmentation. In *ECCV*. 2006.
7. M. Oren, C. Papageorgiou, P. Sinha, E. Osuna, and T. Poggio. Pedestrian detection using wavelet templates. In *IEEE CVPR*, 1997.
8. N. Paragios and R. Deriche. Geodesic active regions for supervised texture segmentation. In *ICCV*. pp. 926-932. 1999.
9. S. D. Pathak, V. Chalana and Y. Kim. Interactive automatic fetal head measurements from ultrasound images using multimedia computer technology. *Ultrasound in Medicine and Biology*, 23 (5), pp 665-673, 1997.
10. M. Rousson, N. Paragios. Shape priors for level set representations. In *ECCV*, 2002.
11. P.J. Schluter, G. Pritchard, and M.A. Gill. Ultrasonic fetal size measurements in Brisbane, Australia. *Australasian Radiology*, 48 (4), pp. 480-486, 2004.
12. Z. Tu. Probabilistic boosting-tree: learning discriminative models for classification, recognition, and clustering. *ICCV*, 2005.
13. P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *IEEE CVPR*, 2001.
14. S. Zhu, and A. Yuille. Region competition: unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. In *IEEE TPAMI*, 18, pp.884-900. 1996.