33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Segmentation of 2D Fetal Ultrasound Images by Exploiting Context Information using Conditional Random Fields

Lalit Gupta, Rajendra Singh Sisodia, V Pallavi, Celine Firtion and Ganesan Ramachandran

Abstract— **This paper proposes a novel approach for segmenting fetal ultrasound images. This problem presents a variety of challenges including high noise, low contrast, and other US imaging properties such as similarity between texture and gray levels of two organs/ tissues. In this paper, we have proposed a Conditional Random Field (CRF) based framework to handle challenges in segmenting fetal ultrasound images. Clinically, it is known that fetus is surrounded by specific maternal tissues, amniotic fluid and placenta. We exploit this context information using CRFs for segmenting the fetal images accurately. The proposed CRF framework uses wavelet based texture features for representing the ultrasound image and Support Vector Machines (SVM) for initial label prediction. Initial results on a limited dataset of real world ultrasound images of fetus are promising. Results show that proposed method could handle the noise and similarity between fetus and its surroundings in ultrasound images.**

I. INTRODUCTION

Ultrasound is widely used by obstetricians as an imaging modality for extracting the biometric and morphological data of fetuses. It plays a key role in dating the pregnancy, detecting anomalies, monitoring the fetal growth *etc.* Key challenges in ultrasound (US) image analysis are noise, low contrast, and other US imaging properties such as similarity between texture and gray levels of two organs/ tissues. Therefore, manual interpretation of ultrasound images and computing biometric data for fetuses is quite a tedious and time consuming task. Moreover, it is also susceptible to human variability. An automation in this area could be helpful in robust diagnosis of fetus and reducing human variability.

In this paper, we address the problem of fetal image segmentation to assist obstetricians for extracting the biometric parameters of fetuses. These parameters are: biparietal diameters (BPD), head circumference (HC), and femur length (FL), and abdominal circumference (AC) [1]. Each of these parameters provides, via a specific mathematical expression, estimation of the gestational age. It is seen that region boundaries in ultrasound images often do not conform to the assumptions of many image processing algorithms due to high noise and variable contrast. In this paper, we handle the problem of noise by using stochastic texture features to represent the image. It is known that fetus is typically surrounded by specific maternal tissues, amniotic fluid and placenta. We use this context information by modeling the image using Conditional Random Fields (CRFs). CRFs also help to handle the similarity between gray values between fetus and its surroundings. By using context information, we also tend to use the shape of the fetus as a feature, which increases the robustness of our algorithm.

In literature, segmentation of ultrasound images are typically application driven. A survey of such approaches is presented by Nobel *et al.* [2]. The application areas they have focused are echocardiography, breast ultrasound, transrectal ultrasound (TRUS), intravascular ultrasound (IVUS) images, and ultrsound images in obstetrics and gynecology. In conclusion, they identified that existing methods for ultrasound image segmentation are not very accurate due to presence of noise. They also concluded that most of the existing methods for ultrasound image segmentation techniques focus on region growing or active contours. These are semi-automatic segmenting systems, in which seed points or initial contours have to be manually identified. The method proposed in this paper handles the noise and also does not require any human intervention for either selecting seed points or initiating contours.

A recently published paper by Shotton *et. al* [3] showed that conditional random fields could be used for exploiting context information from neighboring segments. They outlined a method for appearance, shape and context modeling for multi-class object recognition and segmentation using conditional random fields. In this paper, we propose a combination of texture features, support vector machines and conditional random fields for segmenting fetal ultrasound images. The block diagram of the overall methodology proposed in this paper is given in Fig. 1.

The details of the proposed algorithm along with experimental results has been described in rest of the paper, which is organized as follows: section II introduces the conditional random fields, section III explains the proposed method in detail. It explains the interaction between SVMs and CRFs. Section IV focuses on experimental results and also discusses the advantages of using CRFs with examples. Section V concludes the paper with a discussion on future work.

II. CONDITIONAL RANDOM FIELDS (CRFS)

Kumar and Hebert [4] were the first one to define CRF for image processing applications. They outlined two main differences between the conditional random fields and Markov random fields (MRFs) frameworks as:

1) In CRFs, the association potential at any site is a function of all the observations while in MRFs (with

This work has been done at Philips Reseach Asia - Bangalore
Lalit Gupta (corresponding author) lalit.gupta@philips.com
Rajendra Singh Sisodia rajendra.sisodia@philips.com
V Pallavi pallavi.vajinepalli@philips.com
Celine Firtion celine.firtion@philips.com
Ganesan Ramachandran ganesan.r@philips.com

the assumption of conditional independence of the data), the association potential is a function of data only at that site.
2) The interaction potential/ pairwise potential for each pair of nodes in MRFs is a function of only labels, while in the conditional models it is a function of labels as well as the observations.

Here, first we partition the image and represent it as a graph, where each partition of the image should be viewed as a node in the graph. Each node is connected with its neighbors. The algorithms for image partitioning and neighborhood selection are described in the next section. Assuming the pairwise potentials to be non-zero, the conditional random distribution over all labels Y, given the observation X, can be written as

$$P(Y|X) = \frac{1}{Z} exp(\sum_{i \in S} log(\phi(y_i, x_i)) + \sum_{i \in S} \sum_{j \in N_i} \psi_{ij}) \quad (1)$$

$$\phi_i = \frac{1}{1 + exp(-y_i \frac{d_i}{\tau})} \quad (2)$$

$$\psi_{ij} = y_i y_j h^T g_{ij}(X) \quad (3)$$

Where Z is the normalizing constant, ϕ_i and ψ_{ij} are the association and the interaction potential respectively. The association potential, $\phi(y_i, X)$ can be seen as a measure of how likely a site i will take label y_i given its local features, ignoring the effects of other sites in the image. Interaction potential can be seen as a measure of how labels at the neighboring sites i and j should interact given the observed image X.

We have used support vector machines for computing association potential and initial labels. d_i denotes the output of SVM decision function (Eqn. 4) given a feature vector x_i as input and τ is a constant that can adjust the curve of the logistic function.

$$d_i = \frac{1}{sv} \sum_{j=1}^{sv} (w.x_{ij} - c_i) \quad (4)$$

where sv denotes number of support vectors, w denotes weights and c_i represents the bias.

We have used the log-linear model to define the interaction potential [5], which depends on the inner product of the weight vector h and feature vector $g_{ij}(X)$. The weight vector h is learnt during the training session and $g_{ij}(X)$ is defined as $[1, |x_i - x_j|]^T$, where T is the transpose and $|.|$ represents L1 norm. We have used conditional Maximum Likelihood Estimation (MLE) method [5] for the computation of h.

To infer the region labels corresponding to the segments in image a maximum posterior marginal (MPM) criterion has been used. Each segment/ node is assigned a label that maximizes its margin posterior probability i.e.

$$y_i^* = \underset{y_i \in \{-1,1\}}{\text{argmax}} \ P(y_i|X; h) \quad (5)$$

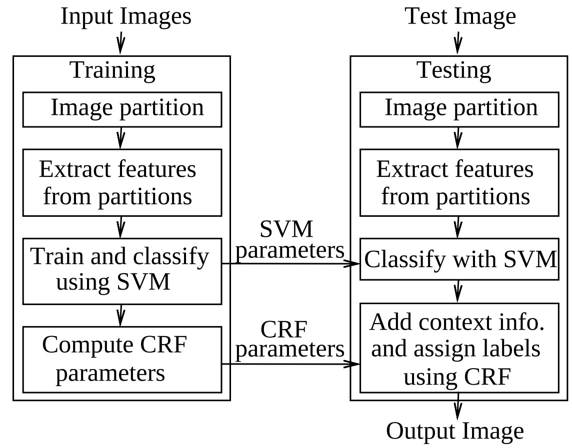
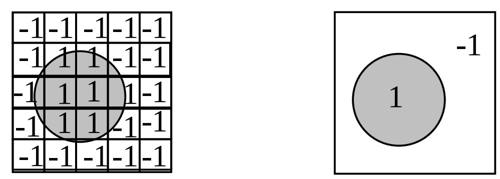


Fig. 1. Block diagram of the proposed methodology.



(a) (b)

Fig. 2. An example of image partitioning: (a) using a grid size of q x r; (b) using texture features and FCM; where the object of interest is shown as a dark circle in the center of the image. '1' represents the foreground and '-1' represents background.

III. Proposed Methodology

The proposed methodology is shown in Fig. 1. There are two phases in our approach, training phase and testing phase. In training phase, all parameters are estimated, which are used during testing. These two phases of algorithms are described further in this section.

A. Training

The images in the training set T are first partitioned into small sub-images or partitions. We have used two different approaches for image partitioning. In the first approach, we divide the image into small sub images (like a grid) of size q x r each. In the second approach, the image is partitioned using a method shown in Fig. 3. The image is clustered using fuzzy-c means (FCM) clustering using texture features. Both of the methods have their advantages and disadvantages. The first method ensures that final result is not dependent on the quality of image partitioning algorithm, hence less error prone, however, it gives blocky boundaries. The second method takes care of the boundaries, however, the features used for partitioning the image should be highly robust to make sure that quality of partitioning is good. A dummy example of two methods is shown in Fig. 2.

The ultrasound images can be seen as stochastic texture images. Therefore, we have extracted Discrete Wavelet Transform (DWT) based texture features from ultrasound images.

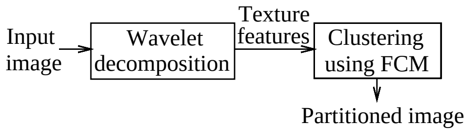


Fig. 3. The methodology for image partitioning using FCM.

The discrete wavelet transform analyzes a signal based on its content in different frequency ranges. Therefore it is very useful in analyzing repetitive patterns such as texture [6]. The 2-D wavelet transform uses a family of wavelet functions and its associated scaling functions to decompose the original image into different subbands, namely the low-low, low-high, high-low and high-high (A, V, H, D respectively) subbands. The decomposition process can be recursively applied to the approximation subband (A) to generate decomposition at the next level.

The filter responses are post-processed to compute the local energy estimates. The absolute value of a filter response $h_l^q(x,y)$ is convolved with a low pass Gaussian post filter $g(x,y)$ to yield a post-filtered energy of the q^{th} subband of l^{th} filter as

$$e_l^q(x,y) = |h_l^q(x,y)| * *g(x,y) \tag{6}$$

The feature vectors computed from the local window around a given pixel from the energy estimates are

1) Mean, $\mu = E[e_l^q(x,y)]$, of post-processed A
2) Variance, $\sigma = E[(e_l^q(x,y) - \mu)^2]$, of post-processed V and H.

Here the $E[.]$ is the expectation operator.

$$x_i = [\mu_A^h(x,y)\sigma_V^h(x,y)\sigma_H^h(x,y)]^T \tag{7}$$

where, x_i is the feature vector, $\mu_A^h(x,y)$ is the estimated mean of the energy in the approximation subband obtained by filtering the input image (using haar wavelet filter), and $\sigma_V^h(x,y)$ is variance of the estimated energy in the vertical subband (using Haar filter).

A three dimension feature vector is obtained by concatenating all features obtained for each of the partition/ sub-image. Hence each partition of the image is now represented by a feature in $\Re^3$. These features are used for SVM based classification.

Once the image is partitioned and features are extracted, each partition is manually assigned as foreground/ background based on the object of interest. The method of manual label assignment is illustrated in Fig. 2. Here we experimented on ultrasound images to extract fetal image. SVM is then trained using this training data. Empirically, we found that a polynomial kernel with degree two is optimal for this problem. The features and training data so obtained is used to compute CRF parameters. The methodology for computing CRF parameters is given in Section 3.

B. Testing

Initially, the test image is partitioned into sub-images. The initial labels of sub-images are assigned using SVM based classification. The SVM parameters computed using training phase are used here for classification. The initial labels are further refined using the CRF model i.e. estimated using the train images. The labels to each sub-image is assigned (foreground or background) using the Eqn. 5.

IV. Experimental Results and Discussion

Our approach requires a learning process to train the model. For this purpose we have used two images. In order to capture the stochastic nature of ultrasound images, we have used texture features i.e. mean of approximation and variance of horizontal and vertical components of Haar filter to model the image. Initial classification is performed using SVM, which is further refined using CRF, which captures the context information. In case of fetal ultrasound image segmentation, it is known that fetus is surrounded by amniotic fluid, which has different texture characteristics to the fetus. CRF uses this context information for segmenting the fetal image accurately.

The proposed methodology has been tested on two fetal images. The test images along with the desired output have been shown in Fig. 4(a)-(b) and Fig. 5(a)-(b). The ground truth is manually drawn with the help of an expert.

The results of segmentation are shown in Fig. 4 - Fig. 5. It can be observed that results from SVM show fetal image to an extent, which is highly refined using CRF.

The results of classification demonstrates the following:

1) The context information i.e. the properties of surrounding tissues/ organs play a key role in fetal image segmentation in ultrasound imaging. It can be observed from results that segmentation has significantly improved using CRF in all the cases.
2) Accurate object identification can be achieved in ultrasound image segmentation without any manual interference using CRFs.
3) The proposed system is robust enough to segment the fetal image accurately even in presence of the interference between the amniotic fluid and the tissues.
4) It can also be observed that using techniques for removing unconnected components instead of CRF does not give desired results as they could remove the object of interest in case it is not connected. And could also include background objects in the cases where they connect with the foreground object. This is apparent from all the results shown.
5) Image partitioning using texture features and FCM gives accurate boundaries in most cases and hence accurate image segmentation as shown in Fig. 5(h). However, as shown in Fig. 4(h) final output has suffered because fetal image region is merged with other tissues in the image due to similar texture characteristics. Manual image partitioning is uniform, therefore quality of segmentation is good in both the cases Fig.

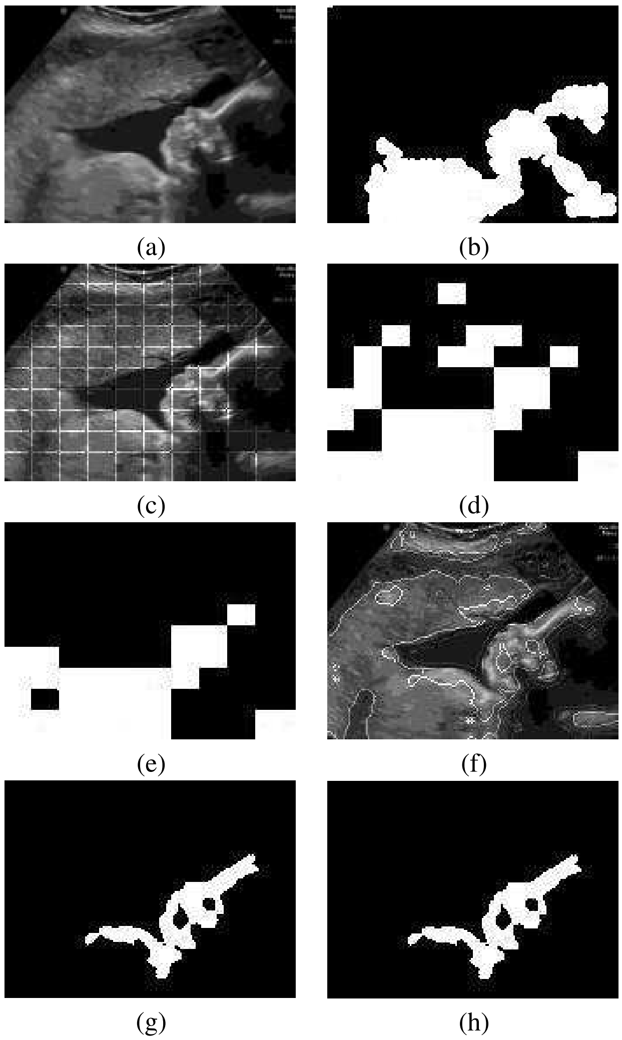


Fig. 4. (a) Fetal ultrasound image; (b) Desired segmented output (marked by an expert); (c) partitioned image (grid lines are superimposed on image for better visualization); (d) output using SVM using partioning as shown in (c); (e) final output after using CRF using partitioning shown in (c); (f) partitioned image using texture features and FCM; (g) output using SVM using partioning in shown in (f); (h) final output after using CRF using partitioning shown in (e).

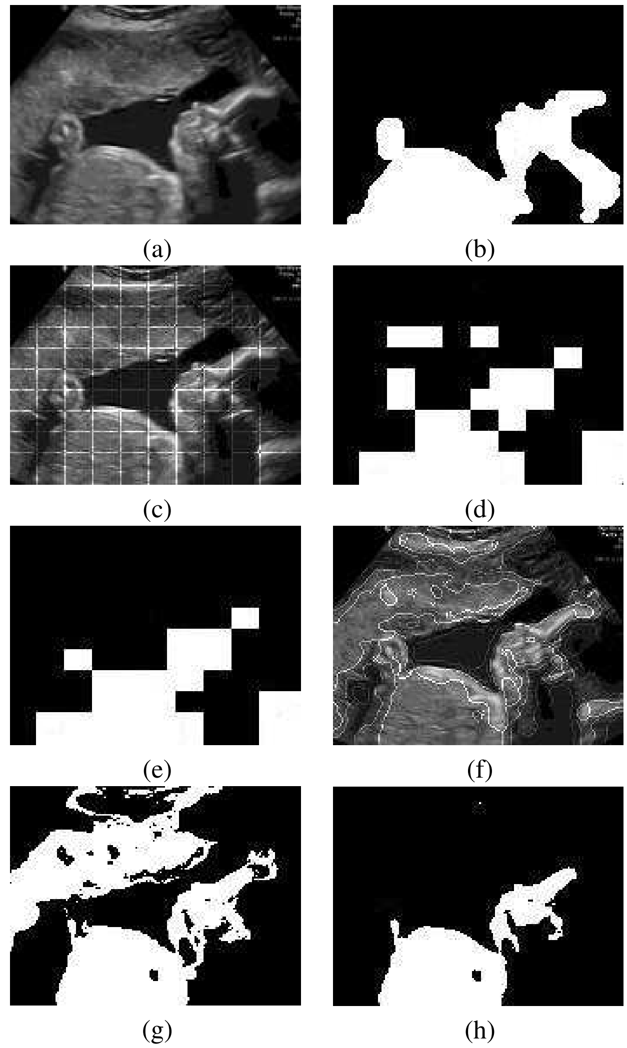


Fig. 5. (a) Fetal ultrasound image; (b) Desired segmented output (marked by an expert); (c) partitioned image; (d) output using SVM using partioning as shown in (c); (e) final output after using CRF using partitioning shown in (c); (f) partitioned image using texture features and FCM; (g) output using SVM using partioning in shown in (f); (h) final output after using CRF using partitioning shown in (e).

4(e) and Fig. 5(e), however, segmentation at boundaries is inaccurate.

V. Conclusion and Future Works

An automated method based on conditional random fields for fetal ultrasound image segmentation is presented in this paper. The method is found to be promising on a limited dataset. The method could further be evaluated on a larger dataset. The results show that context information is an important parameter for segmenting ultrasound sound images. As a scope of future work, various forms of CRF such as Tree CRFs, could be evaluated. An image partitioning method by combining the two presented approaches, could be explored.

References

[1] P. Tolay, V. Pallavi, P. Bhattacharya, C. Firtion, and R. S. Sisodia, "Spine detection in fetal ultrasound images," in *Proceedings of SPIE Medical Imaging*, vol. 7260, Orlando, Florida, USA, 2009.

[2] J. A. Noble and D. Boukerroui, "Ultrasound image segmentation: a survey," *IEEE Transactions on Medical Imaging*, vol. 25, no. 8, pp. 987–1010, 2006.

[3] J. Shotton, J. Winn, C. Rother, and A. Criminisi, "Textonboost: Joing appearance, shape and context modelling for multi-class object recognition and segmentation," *International Journal of Computer Vision*, vol. 81, no. 1, 2009.

[4] S. Kumar and M. Hebert, "Discriminative random fields: A discriminative framework for contextual interaction in classification," in *Proceedings of the Ninth IEEE International Conference on Computer Vision*, vol. 2, 2003, pp. 1150 – 1157.

[5] L. Zhang and Q. Ji, "Image segmentation with a unified graphical model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 8, pp. 1406–1425, 2010.

[6] E. Salari and Z. Ling, "Texture segmentation using hierarchical wavelet decomposition," *Pattern Recognition*, vol. 28, pp. 1819–1824, 1995.