

A Multichannel Markov Random Field Approach for Automated Segmentation of Breast Cancer Tumor in DCE-MRI Data Using Kinetic Observation Model

Ahmed B. Ashraf, Sara Gavenonis, Dania Daye, Carolyn Mies,
Michael Feldman, Mark Rosen, and Despina Kontos

The University of Pennsylvania, Philadelphia, PA 19104, USA
Ahmed.Ashraf@uphs.upenn.edu

Abstract. We present a multichannel extension of Markov random fields (MRFs) for incorporating multiple feature streams in the MRF model. We prove that for making inference queries, any multichannel MRF can be reduced to a single channel MRF provided features in different channels are conditionally independent given the hidden variable. Using this result we incorporate kinetic feature maps derived from breast DCE MRI into the observation model of MRF for tumor segmentation. Our algorithm achieves an ROC AUC of 0.97 for tumor segmentation. We present a comparison against the commonly used approach of fuzzy C-means (FCM) and the more recent method of running FCM on enhancement variance features (FCM-VES). These previous methods give a lower AUC of 0.86 and 0.60 respectively, indicating the superiority of our algorithm. Finally, we investigate the effect of superior segmentation on predicting breast cancer recurrence using kinetic DCE MRI features from the segmented tumor regions. A linear prediction model shows significant prediction improvement when segmenting the tumor using the proposed method, yielding a correlation coefficient $r = 0.78$ ($p < 0.05$) to validated cancer recurrence probabilities, compared to 0.63 and 0.45 when using FCM and FCM-VES respectively.

Keywords: Breast DCE MRI, breast tumor segmentation, tumor characterization, breast cancer recurrence prediction.

1 Introduction

Crucial to the performance of a feature extraction and image classification system is the availability of a reliable segmentation approach for the object of interest (e.g. tumor). In most medical imaging scenarios the automation of this step is particularly important because of the large amount of images to be analyzed. This makes the manual segmentation approach tedious and prohibitively expensive. As a result, a lot of research has been done in the medical imaging community for improving the quality of automated segmentation. Specifically,

there is an abundance of methods in the literature that are geared towards segmenting specific anatomical structures (e.g. parts of the brain). Notable among them are the variants of active contours [1] and active shape models. There are two drawbacks associated with this class of methods. First, they require manual initialization that should be very close to the actual structure to be segmented. Second, they aim at segmenting particular anatomical structures that have well defined control points in their shape, which is often not the case for arbitrarily shaped lesions.

Among the work done on breast tumor segmentation the most popular approach is fuzzy *c*-means (FCM) clustering employed by many researchers due to its simplicity (e.g. [2]). More recently Lee et al. [3] presented a method that runs FCM on the variance map of enhancement kinetics (FCM-VES). These methods do not take into account the overlap of the feature values of the tumor and non-tumor pixels. As a result they have to settle with a manually set threshold on FCM membership probabilities, leading to poor generalization. In order to address these issues we propose to incorporate a component of learning for breast tumor segmentation. We present a multichannel extension of Markov Random Fields (MRFs) to make maximal usage of multiple feature streams derived from the imaging data, here specifically related to kinetic analysis of DCE breast MR images.

In this paper we explore how inference methods like loopy belief propagation [6] may be extended for a multichannel MRF. Zhang et al's work [4] comes closest in concept to our current work. In [4] the authors address the problem of brain image segmentation by making use of conditional independence in MRFs and present an expectation maximization framework to solve the MRF. However, our approach differs fundamentally from [4] in that we exploit conditional independence for solving an MRF via loopy belief propagation [6] that reduces a multichannel MRF to a single channel MRF for inference queries. Moreover, unlike [4], we model our MRF using the superpixel representation that avoids the complexity of dealing with raw pixels. Key contributions of this paper are :

- We prove that for making inference queries, any multi channel MRF can be reduced to a single channel MRF provided the features in different channels are conditionally independent given the hidden variable. (Section 3)
- To elaborate on this premise, we introduce multiple feature channels derived from the kinetic analysis (pixel wise feature maps of enhancement and wash-in-slope) of breast dynamic contrast enhanced (DCE) magnetic resonance (MR) images in the observation model of MRF. (Section 4)
- We show that the segmentation algorithm based on our proposed method yields an AUC of 0.97 under the ROC curve for breast tumor segmentation compared to 0.86 and 0.60 for commonly used previously proposed approaches, [2], [3]. (Section 5)
- Finally we demonstrate that superior segmentation leads to improved breast tumor characterization. We show that the segmentation methods of this paper lead to improvement in predicting validated breast cancer recurrence scores from DCE breast MR images. (Section 6)

We begin with a brief review of single channel Markov random fields.

2 Single Channel MRF

To elaborate on the MRF concept, Figure 1(a) shows how an image can be modeled as a Markov random field. Each node represents the class of a pixel, and neighboring pixels are connected via edges. In the context of segmentation the goal is to infer the class label for each pixel (e.g. foreground vs. background). Throughout the paper we choose the super pixel representation [5] as its computational complexity is much less than the pixel representation. Each node emits an observation which can be any feature e.g. pixel value. The joint probability of super pixel class and feature value over the entire image is given in Eq. 1:

$$P(\mathcal{X}, \mathcal{Y}) \propto \prod_i \phi_i(x_i) \prod_{i,j \in E} \psi_{ij}(x_i, x_j). \quad (1)$$

$$\phi_i(x_i = c) = P(y_i | \mathcal{GMM}_c). \quad (2)$$

$$\psi_{ij}(x_i, x_j) = e^{-\beta \mathcal{I}(x_i \neq x_j)}. \quad (3)$$

In Equation 1, ϕ represents the node-potential and ψ represents the edge potential, and E models the adjacency of nodes including only those nodes that have edges between them. $\phi_i(x_i)$ captures the correlation between the feature and the class label, indicating the likelihood of x_i coming from class c based on the feature value of y_i and the feature distribution of class c . Node potentials can be learned from the training data. In Equation 2, \mathcal{GMM}_c is a Gaussian mixture model learned for class c . In Equation 3, ψ biases neighboring nodes to have the same class label via the parameter β . Approximate MAP (maximum a posteriori) solution to the MRF can be inferred by loopy belief propagation (Loopy BP) [6] which maximizes the joint probability $P(\mathcal{X}, \mathcal{Y})$. Loopy BP is a dynamic message passing algorithm used for doing inference in MRF. $\delta_{i \rightarrow j}(\mathcal{X}_j)$ is defined to be an incoming message into node j from its neighboring node i . To start the process of message passing through the nodes, we need to initialize the node messages (typically to unity) and then the messages can be updated in the next iteration as follows:

$$\delta_{i \rightarrow j}(\mathcal{X}_j) = \frac{1}{Z_{i \rightarrow j}} \sum_i \phi_i(x_i) \cdot \psi_{ij}(x_i, x_j) \prod_{k \in \mathcal{N}(i) - j} \delta_{k \rightarrow i}(x_i). \quad (4)$$

where $Z_{i \rightarrow j}$ is a normalization constant, and $\mathcal{N}(i)$ is a set containing the neighbors of node i . Equation 4 is repeatedly invoked till the messages converge (the update in each message is less than ϵ e.g. 10^{-6}). Once the messages have converged the final inference is done by using Equation 5 below:

$$\hat{P}(\mathcal{X}_i) = \frac{1}{Z_i} \phi_i(x_i) \prod_{k \in \mathcal{N}(i)} \delta_{k \rightarrow i}(\mathcal{X}_i). \quad (5)$$

The inference engine will output for each node a vector of size $C \times 1$ ($C = 2$ for two classes), representing the belief of this node coming from each class. Optimal class label is simply the class with the highest belief.

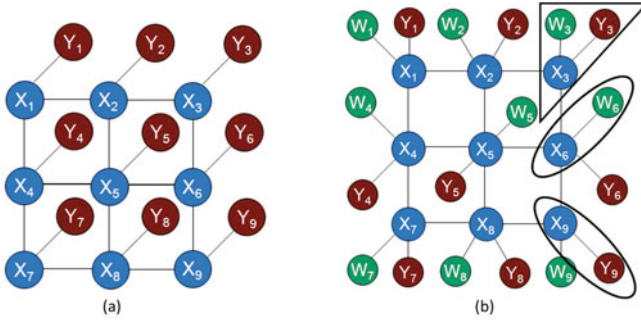


Fig. 1. (a) A single channel MRF: each node represents a super pixel emitting an observation (see Section 2) (b) A multichannel model with two observations . The variables inscribed in the triangle represent the joint distribution of the MRF model. By invoking Theorem 1 we can factorize that joint distribution into two parts represented by the ellipses. (see Section 3)

3 Multichannel MRF

Building on the single channel MRF presented in the last section we now discuss how to incorporate multiple observations in the MRF model. To elaborate on the concept, Figure 1(b) illustrates a multichannel MRF that includes two observations (similar notion could be extended to more than two observations). The joint probability of the super pixel class and the two observations over the entire image is given below:

$$P(\mathcal{X}, \mathcal{Y}, \mathcal{W}) \propto \prod_i \phi_{i(Multichannel)}(x_i, y_i, w_i) \prod_{i,j \in E} \psi_{ij}(x_i, x_j) . \tag{6}$$

$$\phi_{i(Multichannel)}(x_i = c, y_i, w_i) = P(y_i, w_i | \mathcal{GMM}_c) . \tag{7}$$

where $\phi_{i(Multichannel)}$ is the multichannel node potential for the two observations. Our main premise here is that the machinery presented in the previous section can be extended for solving the above multichannel MRF. To this end, we present the following theorem.

Theorem 1: Any multichannel MRF can be reduced to a single channel MRF for making inference queries if the features in different channels are conditionally independent given the hidden (inferred) variable.

Proof: We prove the above theorem for a two channel MRF, while the same notion can be extended to more than two channels without loss of generality. If the features in different channels are conditionally independent given the hidden variable we can factorize the right hand side of Equation 7 as follows:

$$P(y_i, w_i | \mathcal{GMM}_c) = P(y_i | \mathcal{GMM}_c) P(w_i | \mathcal{GMM}_c) \text{ if } Y \perp W | X . \tag{8}$$

Equivalently Equation 8 can be written as:

$$\phi_{i(Multichannel)}(x_i = c, y_i, w_i) = \phi_i^{(y)}(x_i = c, y_i) \cdot \phi_i^{(w)}(x_i = c, w_i) . \tag{9}$$

where $\phi_i^{(y)}$ and $\phi_i^{(w)}$ are the individual node potentials for the features y and w respectively. We can now substitute these individual node potentials in the message update and inference equations as follows:

$$\delta_{i \rightarrow j}(\mathcal{X}_j) = \frac{1}{Z_{i \rightarrow j}} \sum_i \underbrace{\phi_i^{(y)}(x_i = c, y_i) \cdot \phi_i^{(w)}(x_i = c, w_i)} \cdot \psi_{ij}(x_i, x_j) \prod_{k \in \mathcal{N}(i) - j} \delta_{k \rightarrow i}(x_i) .$$

$$\hat{\mathcal{P}}(\mathcal{X}_i) = \frac{1}{Z_i} \underbrace{\phi_i^{(y)}(x_i = c, y_i) \cdot \phi_i^{(w)}(x_i = c, w_i)} \prod_{k \in \mathcal{N}(i)} \delta_{k \rightarrow i}(\mathcal{X}_i) . \quad (10)$$

The individual node potentials in the above equations (the underlined portion) can be pre-computed and be combined as a single potential. As a consequence, the inference engine would now be running queries on a multichannel MRF without incurring any additional cost, effectively reducing it to a single channel MRF, proving the claim of Theorem 1 \square

In the next section we elaborate on the extraction of pixel-wise feature maps to build a kinetic observation model for our multichannel MRF.

4 Kinetic Observation Model

Typically, for DCE MR images we have a pre-contrast image (captured prior to the injection of a contrast agent) and a number of post contrast images, captured at different time points after the injection of the contrast agent. A usual way to quantify the enhancement pattern [3] is to compute percentage enhancements relative to the pre-contrast image. If we compute the relative enhancement on a pixel by pixel basis we can achieve pixel wise maps of the contrast enhancement. For a particular pixel, the enhancement plotted as a function of time is termed as the kinetic curve. In the literature (e.g.[3]) a number of basic features can be computed from this kinetic curve. For example, peak enhancement (PE), time to peak (TTP), wash-in-slope (WIS), wash-out-slope (WOS). Figure 2(a) illustrates

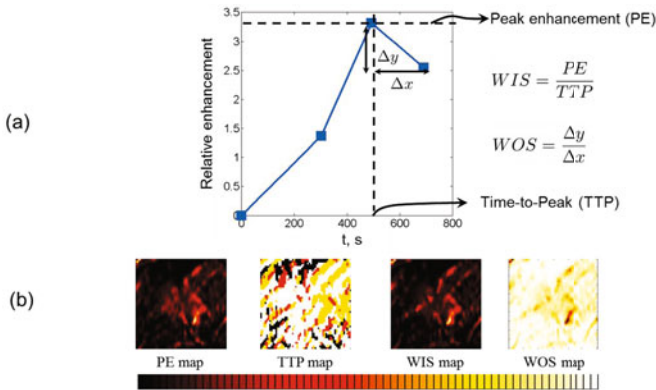


Fig. 2. (a) Illustration of basic kinetic features for a single pixel. (b) Pixel wise maps.

these features for a single pixel. However, we can derive a rich kinetic feature set by computing the pixel-wise map for each feature as depicted in Figure 2(b). We aim to investigate the utility of these feature maps if included as an observation model in our multichannel MRF for the purpose of breast tumor segmentation.

5 Segmentation Experiments

Dataset: All experiments presented in this paper were conducted on DCE breast MR images of 60 women diagnosed with breast cancer. These women had estrogen receptor positive (ER+), node negative tumors, which were analyzed with the Oncotype DX prognostic gene expression assay. Oncotype DX is a validated reverse-transcriptase-polymerase-chain-reaction (RT-PCR) assay (developed by Genomic Health Inc.) that measures the expression of 21 genes in RNA from formalin-fixed paraffin-embedded (FFPE) tumor tissue samples from the primary breast cancer [7]. The final outcome of the Oncotype DX assay is a continuous recurrence score that predicts the likelihood of breast cancer recurrence in 10 years after the treatment. To learn feature statistics for distinguishing the breast tumor and non-tumor area of an image, a fellowship-trained board-certified breast imaging radiologist delineated the lesion boundaries.

Feature selection: To select the features that meet the conditions of Theorem 1 we need to assess the conditional independence of features for which we use the mutual information as a criterion. In the multichannel MRF of Figure 1(b), we are interested in assessing the independence of Y and W given X . Their conditional mutual information is given by:

$$I(Y; W|X) = \sum_{x,y,w} p(x, y, w) \log \frac{p(y, w|x)}{p(y|x)p(w|x)}. \quad (11)$$

We computed $I(Y; W|X)$ by sequentially setting (Y, W) to all possible pairs of the feature maps (post contrast enhancement, peak enhancement, wash-in-slope, wash-out-slope), and retaining the pairs for which $I(Y; W|X) \leq thresh$ (0.02). The feature pair with the minimum $I(Y; W|X)$ was then selected. The above process resulted in the selection of the following two feature maps for our MRF: first post-contrast relative enhancement and the wash-in-slope.

To enable maximal usage of the training data we used a leave-one-out cross validation strategy. To reduce the complexity of the multichannel MRF, we first over-segmented the DCE MR images using the super pixels approach based on normalized graph cuts [5]. In order to define the MRF neighborhood for superpixels we scan the rows and columns of the superpixelized image and look for transitions. This enables to build the adjacency matrix for superpixels that captures the neighborhood. In Figure 3, we show representative segmentation results of our algorithm (row 5). For comparison, we also show segmentation results for FCM as used in [2] (row 3), and results based on running FCM on enhancement variance FCM-VES [3], (row 4). A quantitative comparison in terms of ROC's for the three segmentation strategies is given in Figure 4(a). Our proposed method yields an AUC of 0.97 under the ROC curve as compared to 0.86 and 0.60 for [2] and [3] respectively.

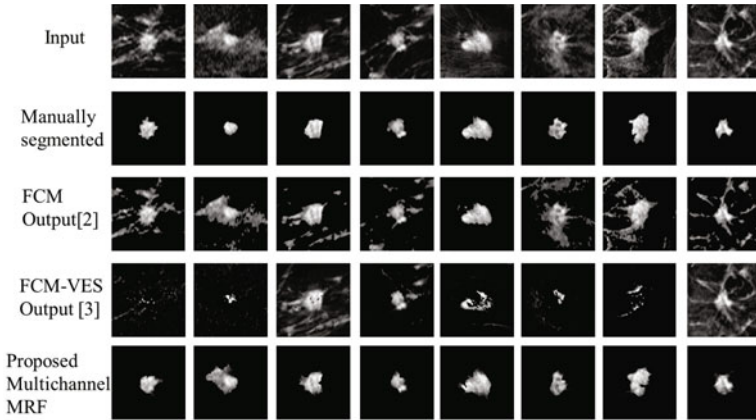


Fig. 3. Qualitative comparison of segmentation results. Results of multichannel MRF approach (Row 5) are qualitatively more similar to the ground truth (Row 2) as compared to traditional approaches (Row 3,4).

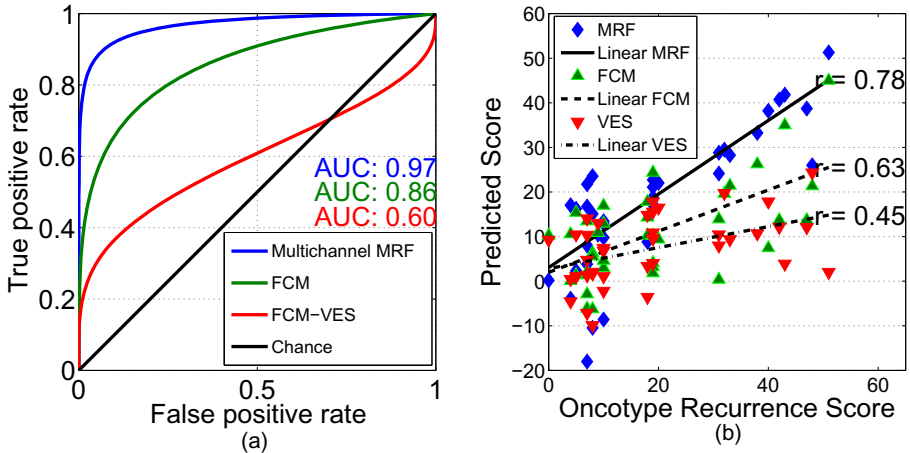


Fig. 4. (a) ROC comparison for breast tumor segmentation. (b) Comparison of Oncotype DX score prediction.

6 Oncotype Score Prediction Experiments

In this section we investigate if better quality segmentation leads to a better tumor characterization. To this end, we aim to predict the Oncotype DX recurrence scores from DCE breast MR images based on previously validated kinetic features [2] extracted within the tumor region segmented as the outcome of our segmentation algorithm. Briefly, in order to extract features from the segmented tumor we cluster the pixels on the basis of their time to peak. In the DCE MR images used in this study there were three post contrast time points. We thus partition the pixels into three groups based on the similarity of their kinetic behavior. Using this partition we first compute the percentage of pixels in each group (this

gives us three features). Within each group we then compute statistics (up to second order) for each of the following maps: peak enhancement, wash-in-slope, wash-out-slope [2],[3]. For every feature map there are three groups and two features per group (mean and variance). In all we build a vector of 21 simple kinetic features. Using these features we learn a linear regression model to predict the continuous Oncotype DX scores. To maximize the utility of the training data we employed a leave-one-out cross validation strategy. The results of the prediction models based on each of the segmentation strategies (multichannel MRF, FCM, FCM-VES) are presented in Figure 4(b). We obtain a correlation coefficient of $r = 0.78$ to the Oncotype DX cancer recurrence scores using our segmentation method as compared to correlations of 0.63 and 0.45 for the other segmentation methods, indicating that our proposed improved segmentation can also result to superior tumor characterization.

7 Concluding Remarks

In this paper we have presented a framework for incorporating multiple feature streams in Markov random fields. We have shown how a multichannel MRF can be reduced effectively to a single channel MRF from the perspective of inference queries. Using this framework we incorporate a kinetic observation model derived from DCE breast MR images into the MRF and demonstrate superior segmentation results as compared to previous methods. Moreover, we also demonstrate that our improved segmentation leads to better tumor characterization. With this framework in place, future research will also investigate the introduction of multi-modality image sequences in the observation model of MRF.

References

1. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. *International Journal of Computer Vision* 1(4), 321–331 (1988)
2. Bhooshan, N., Giger, M.L., Jansen, S.A., Li, H., Lan, L., Newstead, G.M.: Cancerous breast lesions on DCE MR Images: Computerized characterization for Image based prognostic markers. *Radiology* 254(3), 680–690 (2010)
3. Lee, S.H., Kim, J.H., Cho, N., Park, J.S., Yang, Z., Jung, Y.S., Moon, W.K.: Multilevel analysis of spatiotemporal association features for differentiation of tumor enhancement patterns in breast DCE-MRI. *Med. Phys.* 37, 3940 (2010)
4. Zhang, Y., Brady, M., Smith, S.: Segmentation of Brain MR Images Through a Hidden Markov Random Field Model and the Expectation-Maximization Algorithm. *IEEE Trans. Med. Imag.* 20, 45–57 (2001)
5. Ren, X., Malik, J.: Learning a classification model for segmentation. In: *ICCV*, vol. 1, pp. 10–17 (2003)
6. Meltzer, T., Globerson, A., Weiss, Y.: Convergent message passing algorithms – a unifying view. In: *UAI*, pp. 393–401 (2009)
7. Paik, S., Tang, G., Shak, S., Kim, C., Baker, J., Kim, W., Cronin, M., Baehner, F.L., Watson, D., Bryant, J., Costantino, J.P., Geyer, C.E., Wickerham, D.L., Wolmark, N.: Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J. Clin. Oncol.* 24(23), 3726–3734 (2006)