A BOTTOM-UP AND TOP-DOWN MODEL FOR CELL SEGMENTATION USING MULTISPECTRAL DATA

Xuqing Wu and Shishir K. Shah

Department of Computer Science, University of Houston, Houston, TX 77204, USA

ABSTRACT

Cell segmentation is a challenging problem in histology and cytology that can benefit from additional information obtained in using multispectral imaging. Unique transmission spectra of biological tissues are potentially useful for better classification and segmentation of sub-cellular structures. In this paper, we propose a conditional random field (CRF) model that interprets high-dimensional spectral data during inference and pixel labeling. High quality segmentations are computed by combining low-level cues and high-level contextual information extracted by unsupervised topic discovery. Comparative analysis of the proposed model against the commonly used 2-D CRF model in color space is also performed. Results of this evaluation show the benefits of our proposed model.

Index Terms— Conditional Random Fields, Probabilistic Latent Semantic Analysis, Multispectral Image Analysis

1. INTRODUCTION

Up to 7% of the US adult population has single or multiple nodules within the thyroid gland [1] and fine needle aspiration (FNA) is the most cost-effective tool for the initial screening and triage of thyroid nodule cases [2]. In the context of automated analysis of FNA smears, image segmentation plays a crucial role in measuring quantitative parameters of FNA cytology images. The accuracy of nuclei counting, the assessment of cellular heterogeneity and morphometric and karyometric quantification of normal/abnormal cytology are intricately related to the performance of segmentation results. However, manually prepared smears are subject to operator errors causing potential variations in smear thickness, staining times, and imaging parameters, all of which contribute to non-uniform and inconsistent intensity appearance of cells in imaged smears. The importance of cell segmentation in helping computer aided diagnosis is evidenced by the tremendous amount of research devoted to this topic. Nonetheless, accurate and robust segmentation of cells in cytological smears remains a challenge. In this paper, we present a bottom-up and top-down conditional random field (CRF) model that leverages relationships embedded in both spatial and spectral domains for segmenting areas occupied by stained cell nucleus.

CRF is a model-based method used extensively in image segmentation [3, 4]. It directly models the posterior distribution \( p(y|x) \) as a Gibbs field [5] so that the arbitrary dependencies among observations can be captured without any model approximations. Segmentation problems are modeled in CRF with a statistical formulation that imposes spatial regularization by minimizing pairwise and local potentials. CRF is built on the framework of the discriminative approach and it relaxes the strong conditional independence assumption of observations that are typically imposed in traditional Markov Random Field (MRF) frameworks. However, it is difficult to build a CRF model based on low-level cues alone since this bottom-up structure has an untraceable feature space. In recent years, many authors have therefore incorporated contextual information in their models. Ma et al. [6] proposed a coupled CRF model that combines both texture and contour information for categorizing objects. Besbes et al. [7] tried to model contextual interactions by defining superpixels and formulating contextual information in the form of spatial dependencies in the CRF model. Levin et al. [8] approached this problem by using supervised learning on fragments of images. A local search is conducted in this model to find a proper match between local patches and pre-learned fragments. However, the training part of this algorithm is hard to generalize since it can be sensitive to the uniqueness of the geometry of pre-selected fragments. Given a large set of training data, probabilistic Latent Semantic Analysis (pLSA) has been successfully applied to unsupervised topic discovery [9]. This results in efficient representation of contextual information, but fails in effectively capturing spatial relationships. Hence, CRF models can be used along with pLSA to combine low-level cues and high-level contextual information. The combination of these two models has been applied to region classification [10] and natural scene categorization [11].

For a visual discipline as analyzing cytology images, subtle changes in the absorbance spectrum of a stain can only be discerned by more sensitive acquisition techniques [12]. The construction of the color image provides rough maps of the true spectral content of the original sample. Spectral imaging, on the other hand, captures images with accurate spectral content correlated with spatial information and reveals the chemical or anatomic features of the target. In this paper, we present an extension of CRF that models high-dimensional data incorporating spectral constraints dynamically along the neighboring wavelengths. In addition, contextual awareness is embedded in the CRF model with a pLSA representation. The discriminative performance of spectral data is demonstrated by comparing to data in the color space. The key concept of this model is to improve local recognition rate through the unsupervised top-down approach and exploit spatial relationship along spectral domain.

The rest of the paper is organized as follows. Section 2 introduces the proposed CRF framework. Experimental results are presented in section 3 and the paper is concluded in section 4.

2. MODEL REPRESENTATION

2.1. Model description

We consider cell segmentation in spectral images as a classification problem. Each image consists of a number of image patches and each image patch belongs to one of \( C \) classes. Represented by the undirected graphical model, \( G = (V, E) \) with vertices \( V \) and edges \( E \). Each image patch is formulated as a vertex in the graph and each edge connects a pair of image patches. For a sequence of \( M \) images,
let \( x = (x_1, x_2, \ldots, x_M) \), where each element of \( x \) is an observed variable and \( x_i \in \mathbb{Z}^n \). \( n \) is the number of patches in the image. For multispectral data, \( x_i \) is the spectral absorption value acquired at a certain wavelength \( \omega_i \). The corresponding segmentation label is denoted by \( y = (y_1, y_2, \ldots, y_M) \) and \( y_i \in \mathbb{C}^n \). In this paper, the posterior probability distribution of the segmentation \( y \) is modeled by four factors: 1) the local relationship between observed data and the label; 2) the coarse classification information derived from latent topics discovery; 3) spatial constraints along each edge in the 2D image; and 4) spectral constraints imposed by the label state in neighboring spectra.

The energy function of the CRF model, accordingly, consists of a local term \( (E_l) \), a pairwise term \( (E_p) \), and a spectral-constrained term \( (E_s) \).

For a bottom-up model, The local term maps image features extracted from low-level cues to corresponding labels. However, the untraceable feature space makes this mapping a sub-optimal choice. We overcome this limitation by using a coarse top-down approach to reduce the search space through class-specific image information. Each image patch is assigned to a latent topic by using pLSA and the topic is further refined by the local mapping. pLSA is a latent variable model that associates an unobserved class variable \( z_k \in z_1, \ldots, z_K \) with each observation. pLSA uses ‘bag of words’ representation. The data are represented by an observation matrix that indicates the number of occurrence of word \( w \) (rows) in document \( d \) (columns). Translated into image terms, a document is an image. A histogram-based descriptor is extracted from each image patch.

For the \( j \)-th patch in the image \( x_i \), given a local mapping \( g(x_{ij}) \) as a feature extraction function that maps the observation at site \( j \) to a topic, the local term is defined as:

\[
P_l(y_i; x_i) = \sum_j w_j^T \left( \log \frac{g(x_{ij})}{P(g(x_{ij}|d_1, w)} \right)
\]

where \( w \) is the weight that balances the bias between the topic and local appearance. The local mapping \( g \) is constructed by collecting a subset of local patches of each topic and fitting a Gaussian model to it.

The pairwise term assumes that the same label should be assigned to image patches in the neighborhood. But a penalty needs to be imposed if the assignment of labels is not aligned with image edges. The gradient penalty function is designed to meet both objectives given the strength of the edge as \( \nabla \):

\[
q(\nabla) = \exp \left( -\sum_{k=2}^{k} \right)
\]

where parameter \( k \) controls the penalty decrease/increase rate according to the edge strength. In this paper, we set \( k = 0.5 \). Thus, the pairwise term is expressed as:

\[
E_p(y_i; x_i) = \sum_{j,k} q(\nabla_{jk}) |y_{ij} - y_{ik}|
\]

Minimizing \( E_p(i) \) means that it is cheaper if labeling discontinuities are aligned with image discontinuities.

Figure 1(a) shows the graphical model of CRF involving the local term and pairwise term. The label on any site is related to the measurement of local features as well as influenced by neighboring sites in the 2D image. For the multispectral images in the 3D space, we can model them independently along the spectral dimension. However, the independence assumption is against the physical evidence of strong correlation that exists among images acquired at different wavelengths. Similar to the random fields model in the 2D image, we assume that images in the spectral field satisfy the local Markov property, where, for a sequence of spectral image, the distribution of an image depends only on images in neighboring wavelength fields. In other words, for any image patch \( j \) in the \( \omega \)-th frame (wavelength = \( \omega \)) in a spectral image, its label is related to its spectral neighbors \( y_i-1, N(j)+j \) and \( y_i+1, N(j)+j \), where \( N(j) + j \) refers to image patch \( j \) and its neighbors.

Using the Bayes’ rule, the spectral constrained model can be expressed as:

\[
p(y_i, j \mid y_i-1, N(j)+j, y_i+1, N(j)+j, X) = p(y_i, j \mid y_i-1, N(j)+j, y_i+1, N(j)+j, X) * p(y_i-1, N(j)+j, y_i+1, N(j)+j, X)
\]

If we model labels at different wavelength as being conditionally dependent on observed data and labels of spectral neighbors, the equation (6) indicates the following relationship:

\[
p(y_i, j \mid y_i-1, N(j)+j, y_i+1, N(j)+j, x_i) \propto p(y_i, j \mid y_i-1, N(j)+j, y_i+1, N(j)+j, x_i)
\]

Figure 1(b) shows the corresponding graphical model after combining local, 2D spatial, and spectral constraints together. In this new CRF model, the label of an image patch is not only modeled based on the local and spatial constraints, but also influenced by the labels of corresponding image patches in the neighboring wavelength fields. The spectral constraints encourage contiguous labeling in the wavelength neighborhood and are subject to variations of the absorption ratio of different biological components under different wavelengths. Thus, the spectral constraints should be aligned with spectral profile gradient, which is similar to the intensity gradient. The spectral profile is created in two steps: first calculate the mean intensity value of each patch and its neighbors, then plot the mean value across the wavelength field. Let \( f(\omega) \) be the function of the spectral profile in the continuous space, the spectral profile gradient \( s(\nabla) \) at wavelength \( \omega \) is defined as \( \frac{d}{d\omega} f(\omega) \). Similar to the spatial constraint defined in equation (5), we model the spectral-constrained term as:

\[
E_s(y_i; x_i) = \sum_j \left( \frac{1}{s(\nabla_{ij})} |y_{ij-1, N(j)+j} - y_{ij}| + \frac{1}{s(\nabla_{i+1, j})} |y_{ij} - y_{i+1, N(j)+j}| \right)
\]

If the image with wavelength of \( \omega_{i-1} \) or \( \omega_{i+1} \) does not exist, then the corresponding term in equation (8) is set to be zero. Thus, only weak correlation exists between labels of two neighboring image patches if the spectral profile gradient is large.
Combining the local term, pairwise term and spectral-constrained term, CRF energy function is:

\[ E(y; x, \theta) = \sum_i \lambda E_l(y_i; x_i) + \beta E_p(y_i; x_i) + \gamma E_s(y_i; x_i) \quad (9) \]

where \((\lambda, \beta, \gamma)\) are the weights and \(\theta = \{w, \lambda, \beta, \gamma\}\) are parameters that need to be learned through training. Using the energy definition in (9), the probability of labels \(y\) conditioned on observations is:

\[ p(y|x) = \frac{1}{Z} e^{-E(y;x)} \quad (10) \]

where \(Z\) is a normalization term and \(Z = \sum_y e^{-E(y;x)}\).

Fig. 1. CRF model with local, spatial and spectral constraints. In (b), we only show spectral constraints of \(y_{i,1}\) with its \((i+1)\) wavelength field neighbor.

2.2. Parameter estimation and inference

Computing the global minimum for equation (10) is NP-hard [14]. Thus, we use approximation techniques for model learning and inference. Szummer et al. [15] proposed a maximum-margin learning method for parameters estimation based on graph-cut algorithm. For a pre-segmented set of training data \((y^{(n)}, x^{(n)})\), the maximum-margin learning strategy can be represented as a standard quadratic problem:

\[ \min_{\theta} \frac{1}{2} \|\theta\|^2 \text{ s.t.} \]

\[ E(y; x^{(n)}, \theta) - E(y^{(n)}; x^{(n)}, \theta) \geq 1 \forall y \neq y^{(n)} \forall n \quad (11) \]

Given the training data set and parameters \(\theta\), \(E(y^{(n)}; x^{(n)}, \theta)\) can be calculated through equation (9). To get the value of \(E(y; x^{(n)}, \theta)\), we have to find the set of labels that have the minimum energy first:

\[ \arg\min_y E(y; x^{(n)}, \theta) \quad (12) \]

Alpha-expansion graph-cuts [14] is chosen to solve the problem in (12). The algorithm for the model learning is summarized in algorithm 1.

3. EXPERIMENTAL RESULTS

Images to be segmented are obtained from slides of cytological smears that are Papanicolaou stained for the purpose of differentiating pathologies in fine needle aspirated cells from thyroid nodules. In this paper, we use the Olympus BX51 optical microscope and a grating based spectral light source. 2D images are acquired by using a high resolution CCD camera. The Czerny-Turner type monochromator from PTI can provide a tunable light emission spectrum at 10nm resolution. A wavelength range from 400nm - 700nm is used in this study. A total number of 31 pictures are taken with wavelength separation of 10nm. The images are acquired by using the Photometric SenSys™CCD camera having 768 x 512 pixels (9x9μm) at 8-bit digitization. The condenser, aperture diaphragm, and the field stop were kept constant during measurements. Photometric calibration is done automatically by an image similarity based method [16].

The goal of these experiments is to evaluate the discriminative power of spectral data versus data in the color space for segmenting areas occupied by stained cell nucleus. Compared to the extended CRF model formulated for spectral data, CRF model applied to the segmentation of color image does not have the spectral-constrained term. Samples for the test are taken from 12 smears with total of 46 images. For each smear, we annotated several spots in different locations with equal number of cell and background regions. Both color and multispectral images are acquired for each location and there was no overlap between data sets. Test and training were performed in a smear-specific manner. Half of the annotated spots were used to train the parameters and the other half (disjoint data sets) were reserved for testing. To quantify the segmentation performance, the segmentation accuracy is measured by the global consistency error (GCE), a well-established standard measurement used in image segmentation problems [17]. In the following context, all error estimations of the segmentation are referred as the GCE. Images segmented manually are used as the ground truth.

To compare the performance of the segmentation using spectral images against results using color images, we measured both accuracy of segmentations and running time for labeling. Performance measurements are listed in table 1. Tests on all 46 images show that the average segmentation accuracy increases to 85.3% when applying the proposed CRF model to spectral images compared to the 79.6% accuracy rate with color images. The average running time for the segmentation process is 101.74 secs with spectral data and 20.54 secs with color data. This indicates a 5 fold slow down, but the number of images processed with each spectral image is 31 compared to 3 for color images. Figure 2 shows a sample image...
and ground truth of the segmentation. In figure 3 we present segmentation results based on spectral data (figure 3(a)) and color data (figure 3(b)). Figure 4(a) indicates the accuracy score across all samples. Points in the upper-left part of the plot denote that segmentation performs better with spectral data. Figure 4(b) shows the variance in segmentation accuracy across the test set. As seen, the segmentation based on spectral data delivers more consistent results with reduced variance.

Table 1. Segmentation accuracy and running time for labeling.

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
<th>Running time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>spectral data</td>
<td>85.3%</td>
<td>101.74</td>
</tr>
<tr>
<td>color data</td>
<td>79.6%</td>
<td>20.54</td>
</tr>
</tbody>
</table>

Fig. 2. Sample color images of thyroid FNA cytological smears.

(a) spectral segmentation  (b) color segmentation

Fig. 3. Segmentation results using the standard CRF model and the extended CRF model on two example images.

4. CONCLUSION

In this paper, a CRF based model is proposed to do inference with high dimensional spectral information. Relationship of data sets between neighboring spectral fields are weighted by parameters derived from spectral profiles of regions to be segmented and constraints between neighboring wavelength fields are dynamically updated during the process. Statistical inference is executed by using graph-cuts method. The model is trained through a maximum-margin learning strategy. Compared to the traditional maximum likelihood setting, the normalization term $Z$ is canceled in the new learning objective. As a result, the inference in the learning process becomes easier and faster. Testing results with cytological samples show that spatial and spectral constraints defined in the model are beneficial in situations when the image contrast is low and noise-to-signal ratio is high. Spectral data, in general, demonstrates more discriminative power than color data for the problem of cell segmentation in cytological smears.

5. REFERENCES