LESION DETECTION AND SEGMENTATION IN UTERINE CERVIX IMAGES USING AN ARC-LEVEL MRF

Amir Alush, Hayit Greenspan

Bio-medical Engineering, Tel-Aviv University

Jacob Goldberger

School of Engineering, Bar-Ilan University

ABSTRACT

This study develops a procedure for automatic extraction and segmentation of a class-specific object (or region) by learning class-specific boundaries. We present and evaluate the method with a specific focus on the detection of lesion regions in uterine cervix images. The watershed map of the input image is modeled using MRF in which watershed regions correspond to binary random variables indicating whether the region is part of the lesion tissue or not. The local pairwise factors on the arcs of the watershed map indicate whether the arc is part of the object boundary. The factors are based on supervised learning of a visual word distribution. Final lesion region segmentation is obtained using a loopy belief propagation applied to the watershed arc-level MRF. Experimental results on real data show state-of-the-art segmentation results in this very challenging task. If needed, the results can be interactively even improved.

Index Terms— lesion segmentation, MRF, loopy BP.

1. INTRODUCTION

This work is motivated by the need to automatically segment lesion regions in uterine-cervix images (otherwise termed "Cervigrams"). The National Cancer Institute (NCI), National Institute of Health (NIH), has collected 100,000 cervigrams [2]. This vast amount of data requires automated analysis tools as means of detection, diagnosis and cervical cancer research. The automated extraction and analysis of Cervigram tissues is a very complex and challenging task (see image examples throughout this paper). The tissues contain complex and confusing information, they are represented via a narrow dynamic range of colors, and the boundaries between them are not always clear. The Cervigram acquisition process generates additional analysis challenges. Due to the strong flash of the camera and convex shape of the cervix the image tends to be brighter around the cervix center and the illumination decreases gradually towards the cervix border. This results in an inhomogeneous appearance within and across the tissues which automatic segmentation algorithms fail to process correctly. Bright regions may be misclassified as AW lesions, while AW lesions, located in the shaded regions, may not be detected. Additional artifacts that interfere with the tissue segmentation comes from the specular reflections (SR). These artifacts are small and bright regions on the cervix surface, which are generated during the image acquisition process due to the presence of fluids. Finally, the large diversity of cervix shapes, the uneven surface of the cervix and the unconstrained alignment between the cervix center and the position of the camera, introduce extensive variability in the intensity and shape of the cervix, across the image archive. Previous work on the analysis of Cervigram images has focused on automated landmark extraction, including the extraction of the cervix boundary, detection of the Os, and detection (and elimination) of specular reflections [3]. The task of tissue segmentation, and in particular, the clinically important task of AW lesion detection, remains, as yet, unsolved. Initial attempts have been made at tissue segmentation within the cervix region, using pixel-level (region-based) features for the classification task. The studies to date usually focus on one specific analysis task (e.g. single landmark or tissue) or report initial results, with a small number of image examples. The large tissue overlap in feature space has hindered such attempts. Moreover, pixel based classification results in numerous fragmented regions, many of which are false-positives.

In the current methodology, we shift from region-based classification approaches to a boundary-based approach. The boundary is an important key in human expert lesion segmentation and contains a great amount of unexplored information. It can provide strong evidence for the presence of a lesion region, in particular in cases in which the lesion tissue characteristics and surrounding tissue characteristics are non-distinct (i.e. brightness, color and texture cues show considerable overlap). This is the case in the AW lesion segmentation task which is the focus of our current work. Note that this is a common underlying difficulty in other challenging tasks involving tissue segmentation and lesion segmentation in general.

2. LEARNING A PROBABILISTIC EDGE MAP

Our object detection and segmentation approach is based on supervised learning of the object contour using labeled im-

H. Greenspan is currently on Sabbatical at IBM Almaden Research Center, San Jose, CA 95120.

(a) (b) (c)

Fig. 1. MRF construction: (a) Input image cropped to region of interest. (b) The watershed map overlayed on image. (c) The arc-level MRF. Ground truth lesion delineation shown in gray.

ages. Each image is first over-segmented into superpixels by applying the Watershed transform[7] to the color gradient image. The generated segments possess coherent region features and their boundaries are aligned with the image gradients. We refer to this representation as the "Watershed edge map", the generated segments are referred to as superpixels and an arc is defined as the curved line segment between two adjacent superpixels. It is reasonable to expect that watershed map boundaries overlap or are only slightly misaligned with the object's true boundaries [4]. As a result of this step only watershed edge pixels are considered candidate members of the object boundary. The next step is feature extraction for each pixel in the watershed edge map. Patches of size $n \times n$ (we used n=11) are extracted for each pixel on the watershed edge map and are rotated such that the watershed edge line passes horizontally through the patch's center. This makes the features rotationally invariant up to a flip factor [5]. Note that color is not taken into account. The patches are then represented as one dimensional vectors of size n^2. Each vector is normalized by subtracting its mean. The normalization step further increases the algorithm's robustness by making the features invariant to gray-level scale differences. To reduce both the algorithm's computational complexity and the level of noise we apply a principal component analysis procedure (PCA) to reduce the dimensionality of the data. The K-means algorithm is used for building a dictionary, providing data vectors in the projected space that are clustered into M groups (we used M=100). Finally, the centroid of each group is taken to form a dictionary with M visual words. Note that this dictionary learning step is done in an unsupervised mode without any reference to the label of each patch.

Based on the labeled training set, object's boundary and non-object-boundary edge pixels are statistically modeled as frequency occurrence histograms of the dictionary words. We take the same patches that were used to compile the visual dictionary. We assign each patch to the nearest dictionary word (using the Euclidian distance). Since in the training images the object boundaries are given, we have a binary label (boundary/non-boundary) for each patch on the watershed map. If the object boundary does not fall exactly on the watershed edge map we label the nearest edge map as a boundary pixel. We next build two word frequency histograms, one for boundary pixels and one for non-boundary pixels. The first histogram represents the number of times every word from the dictionary is used in watershed edge pixels that are part of the object boundary and the second histogram is similarly defined. Normalizing the histograms we can view them as discrete distributions $P_{\text{boundary}}(\cdot)$ and $P_{\text{non-boundary}}(\cdot)$ of the visual words in the object boundary and non-boundary watershed pixels.

3. THE SUPERPIXEL LEVEL MRF

Given a new image, our goal is to detect and segment the object of interest. First, each one of the watershed edge pixels is translated into one of the visual words from the dictionary. This is done by normalizing the patch vector centered at the edge pixel, rotating it, and applying the PCA transformation that was learned in the training step. Then, every transformed vector is assigned to its nearest word from the dictionary (based on the Euclidian distance). Assuming that the object boundary is a part of the watershed map, either all the pixels in a given arc are on the boundary or none of them are. Hence we need to convert the local pixel-level probabilities into arc-level ones.

There are several ways to transform the probabilities of the pixels in a given arc into a single arc-level probability. We define the probability of an arc to be a part of the object boundary as the average of all the probabilities of the pixels on that arc. The mathematical interpretation of this averaging is based on considering each pixel on the arc as evidence of the boundary/no-boundary attribute of the arc, based on independently sampled noise. Formally, the arc-level probability is:

$$p(\text{arc}|\text{boundary}) = \frac{1}{|\text{arc}|} \sum_{u} p_{\text{boundary}}(u) \qquad (1)$$

where the sum is over all the pixels on the arc. We define $p(\text{arc}|\text{no-boundary})$ in a similar way.

For an arc between adjacent superpixel i and superpixel j we use the following notation for the probability of the observed arc:

$$\phi_{ij}(x_i, x_j) = \begin{cases} p(\text{arc}|\text{no-boundary}) & \text{if } x_i = x_j \\ p(\text{arc}|\text{boundary}) & \text{if } x_i \neq x_j \end{cases} \qquad (2)$$

where x_i and x_j are binary variables such that $x_i = 1$ means that the i-th superpixel is part of the object and $x_i = 0$ means that i-th superpixel is outside the object.

Next we take the information provided by the arcs one step further towards object segmentation. Our goal is to segment the image into two distinct labels of foreground (object) and background. We use the separating arcs as an indictor to whether two adjacent superpixels have the same label. An arc with a strong probability of being a part of the

object boundary is expected to exist between neighboring superpixels with different labels. To translate this intuition into a rigorous mathematical model we view the watershed map as a probabilistic graphical model i.e. a MRF with an irregular grid of superpixels. Each arc contributes a factor. Note that we only use edge information and not region-based information. The fact that the superpixels "content" is not taken into consideration, can be advantageous when we cannot differentiate between object regions and non-object regions as in the case in Cervigram images. Nevertheless, if there is region based information it can be easily incorporated into the MRF as single-variable factors.

Let $x_1, ..., x_n$ be a set of jointly distributed binary random variables associated with the watershed superpixels. From the previous step of the algorithm we obtain local information on each arc. Let $\phi_{ij}(x_i, x_j)$, defined in Eq. (2), be the pairwise factor of the i-th and j-th superpixels. The joint probability function is:

$$p(x_1, ..., x_n) = \frac{1}{Z} \prod_{\{i,j\} \in E} \phi_{ij}(x_i, x_j) \qquad (3)$$

where E is the set of all the arcs in the watershed map. The preferred configuration is one that guarantees that adjacent superpixels with the same label will not be separated by an arc with a strong probability of being part of the object boundary. Hence, the preferred segmentation is the most likely state configuration:

$$\hat{x} = \arg\max_x p(x) \qquad (4)$$

The marginal probabilities of the MRF variables (superpixels) correspond to the posterior probabilities to be a part of the object. Since the graph is loopy, it is not feasible to compute the exact marginal. Instead, we utilize the belief propagation (BP) algorithm. The messages of the loopy belief propagation are as follows. The message from superpixel i to a neighbour superpixel j is:

$$m_{i \to j}(x_j) = \sum_{x_i} \phi_{ij}(x_i, x_j) \prod_{k \in N(i) \setminus j} m_{k \to i}(x_i) \qquad (5)$$

where $N(i)$ is the set of all the superpixels that share a common arc with superpixel i. The approximate marginal distribution of x_i, the belief of x_i, is:

$$\text{belief}_i(x_i) = \prod_{k \in N(i)} m_{k \to i}(x_i) \qquad (6)$$

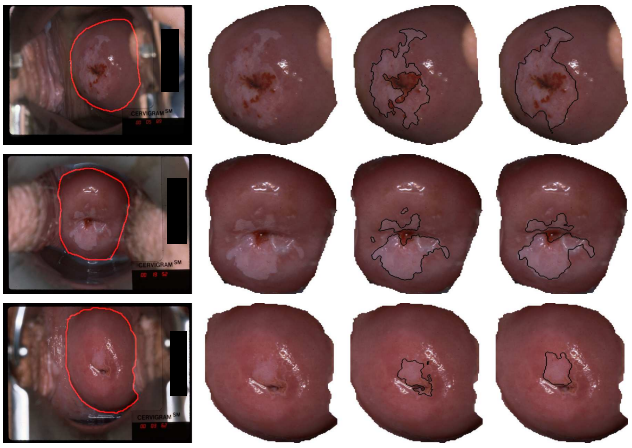
We iterate until the messages converge or until a predefined number of iterations. As a last step we threshold the belief to obtain a hard-decision label for each superpixel. Fig. 1 shows an example of the watershed MRF representation.

4. INCORPORATING USER MARKERS

So far we described a completely automated segmentation procedure which can produce state-of-the-art results for the task of AW segmentation in cervigrams. In this section we extend the algorithm to incorporate user interaction. As shown in Fig. 3, in some cases the automatic segmentation is either lacking or erroneous especially in parts of the image that resemble the AW; this in one of the main challenges in this task. In such cases a user knowledge can be utilized to improve the segmentation by receiving user markers for object and/or background. The markers can be entered either by mouse strokes or mouse clicks. These markers are interpreted by the system as a user indication on the true label of the (hidden to the user) watershed superpixels that contain the markers. The only change in the arc-level BP algorithm is related to the messages from marked superpixels which now take the following form:

$$m_{i \to j}(x_j) = \phi_{ij}(x_i, x_j) \quad \text{if } x_i \text{ is given} \qquad (7)$$

where x_i is the object/bacground information provided by the user on (markers that are placed in) superpixel i. All other messages related to superpixels that were not marked by the user remain as they are defined in Eq. (5). Also the final 'belief' of superpixels marked by the user is set to be the value specified by the user instead of using Eq. (6).



(a) (b) (c) (d)

Fig. 2. Automatic segmentation results: (a) Input image. Region of interest marked in red. (b) Region of interest enlarged. (c) Our segmentation (black line). (d) Expert's segmentation (black line).

5. EXPERIMENTAL RESULTS

Experimental results are shown on a dataset of Cervigram images provided by the National Cancer Institute and the National Library of Medicine at NIH. Our system was trained using 11 randomly selected images from the Cervigram dataset; the remainder of the images (200) was used for testing. The data set also contained two experts annotated ground truth

Table 1. Comparative segmentation results

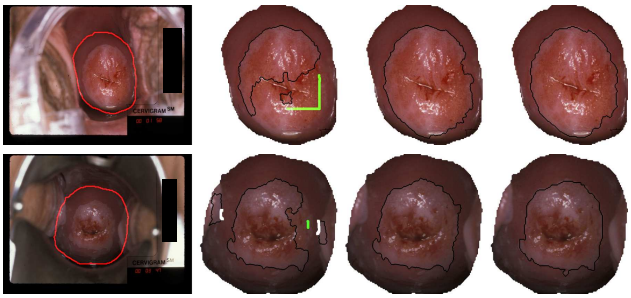
	Rand	False-positive	Dice
BEL	0.6	0.36	0.46
Automatic arc-level BP	0.82	0.04	0.62
Interactive arc-level BP	0.85	0.03	0.70

boundaries for the AW lesion region. Both training and testing were down-sampled to a size of approximately 500×500 pixels, and only the Cervix boundary, marked by an expert, was considered in the analysis (see Fig. 2).

Automatic segmentation results are shown in Fig. 2. Note the large similarity between the extracted lesion region (c) and the GT (d). Note also that the expert tends to mark an encompassing elliptical region of the lesion, which includes additional detail, such as the Os and the CE tissue surrounding it (in dark red). The automated segmentation algorithm removes this detail from the delineated region. Since the object segmentation is performed on superpixels instead of pixels, the algorithm is very efficient. It takes less than a second to detect and segment the lesion region in a Cervigram image. Times were measured on dual quad-core Intel Xeon 2.33 GHz.

We next evaluated systematically the performance of the automated lesion region segmentation results of the proposed algorithm. The comparison is with a region based classification variant of the BEL algorithm, which was trained using pixels on/off the object as true/false examples [1, 8]. The BEL algorithm optimal parameters are used. Both algorithms are trained/tested on the same images.We use the following standard measures: Rand Index [6], Dice measure and False Positives (FP) measure. A statistical summary of the results averaged for the 200 test image using both expert's annotations is presented in Table 1. The proposed scheme compares favorably with the state-of-the-art BEL algorithm.

In order to verify that the interactive segmentation can improve the automated segmentation when correct markers are entered, we have performed a quantitative evaluation on our entire test set using a simulation that found automatically the superpixels of disagreement for both background and foreground between the GT and the automated segmentation result. It then randomly chose from each image ten such superpixels, if present, "marked" foreground/background markers on them and performed the segmentation as described in Section 4. A statistical summary of the interactive segmentation results is presented in Table 1. We also show several examples of interactive segmentation performed manually using mouse strokes markers are shown in Fig. 3. Note how for the upper cervigram the manual markers helped segment the lower part of the AW which was missed by the automatic segmentation. In the lower cervigram the markers helped excluding from the automated segmentation the vaginal walls which resembles the AW.



(a) (b) (c) (d)

Fig. 3. Interactive segmentation results: (a) Input image. Region of interest marked in red. (b) Automatic segmentation (black line). User markers are colored: green (object), white (background). (c) The final segmentation result (black line). (d) Expert's segmentation (black line).

To conclude, in this study we presented an automatic and interactive segmentation methodology for lesion regions within uterine cervix images. As far as we know, this is the first large-scale work ever to be published on extracting lesion regions automatically in Cervigram images. The method presented is a general one. We are currently testing additional tasks such as lesion segmentation in the liver and brain.

6. REFERENCES

[1] P. Dollar, Z. Tu, and S. Belongie. Supervised learning of edges and object boundaries. *Computer Vision and Pattern Recognition (CVPR)*, 2006.

[2] R. Herrero et al. Population-based study of human papillomavirus infection and cervical neoplasia in rural Costa Rica. *Journal of the National Cancer Institute*, pages 464–474, 2000.

[3] H. Greenspan, S. Gordon, G. Zimmerman, S. Lotenberg, J. Jeronimo, S. Antani, and R. Long. Automatic detection of anatomical landmarks in uterine cervix images. *IEEE Trans. on Medical Imaging*, pages 454–468, 2009.

[4] Y. Li, J. Sun, C.-K Tang, and H.Y. Shum. Lazy snapping. *ACM SIGGRAPH*, pages 303–308, 2004.

[5] M. Prasad, A. Zisserman, A. W. Fitzgibbon, M. P. Kumar, and P. H. S. Torr. Learning class-specific edges for object detection and segmentation. *Indian Conference on Computer Vision, Graphics and Image Processing*, 2006.

[6] W. M. Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, pages 846–850, 1971.

[7] P. Soille. Morphological image analysis. principles and applications. *DCGI*, 2003.

[8] S. Zheng, Z. Tu, and A.L. Yuille. Detecting object boundaries using low-, mid-, and high-level information. *Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2007.