

Automated and Interactive Lesion Detection and Segmentation in Uterine Cervix Images

Amir Alush, Hayit Greenspan, Jacob Goldberger

Abstract—This paper presents a procedure for automatic extraction and segmentation of a class-specific object (or region) by learning class-specific boundaries. We describe and evaluate the method with a specific focus on the detection of lesion regions in uterine cervix images. The watershed segmentation map of the input image is modeled using an 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. The 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 on this very challenging task that if necessary, can be interactively enhanced.

Index Terms—MRF, watershed map, uterine cervix, cervigrams, visual words, lesion segmentation, lesion detection.

I. INTRODUCTION

Cervical cancer, one of the most common cancers affecting women worldwide and the most common in developing countries (Parkin *et al.* [28], Eluf-Neto and Nascimento [7]) can be cured in almost all patients, if detected and treated in time. However, the incidence of cervical cancer and mortality rates remain high in resource-poor regions, where high quality Pap (cytology) screening programs often cannot be sustained because of inherent complexity and cost. An alternative method of cervical cancer screening, known as visual inspection with acetic acid, is based on the color change of cervix tissues when exposed to acetic acid. This inexpensive method helps to detect abnormal cells that turn white (acetowhite) following the application of 3%-5% acetic acid (Wright [19]). An analogous photographic method that permits archiving and study is cervicography. Cervicography was first described by Staff in 1981 [39]. In this method the uterine cervix is photographed with a special fixed-focus 35mm camera equipped with a ring flash used to provide enhanced illumination of the target region. Fig. 1 shows an example of a cervicographic image. During the image acquisition process the photographer manually moves the camera back and forth to get the image in focus. The fixed focus of the camera preserves a constant

distance between the camera and the cervix, which makes it possible to have comparable pictures of the cervixes of all patients and to measure areas within the cervix. Immediately before the pictures are taken the cervix is washed with 3%-5% acetic acid for one minute. The acetic acid facilitates the removal of any remaining mucus and highlights any abnormal epithelium.

Later, an expert in cervical pathology projects these pictures onto a screen to obtain a magnified image of the cervix and inspects the infected or precancerous epithelium. Visual methods like cervicography may be used prior to colposcopic and/or Pap smear screening if available on patients with indicators of concern. If required the patient is then referred for medical treatment.

The National Cancer Institute (NCI) and the National Institutes of Health (NIH) have collected a substantial amount of biomedical information related to the occurrence and evolution of uterine cervical cancer in longitudinal multi-year studies carried out in Guanacaste, Costa Rica, and in the United States. 100,000 cervicographic images in the form of 35mm color slides, as well as medical classifications for the cervigrams into diagnostic categories were collected (Herrero *et al.* [14], [15], Schiffman and Castle [35]).

A major long-term objective is to develop a unique Web-accessible database of digitized cervix images to investigate the role of human papillomavirus (HPV) that infects the epidermis and mucous membranes of humans in the development of cervical cancer and its intraepithelial precursor lesions in women [25]. The cervigrams in the NIH archive are unlabeled and have no attached annotation that describes their visual features. Manual extraction of important regions within the cervix by medical experts is not feasible due to the enormous number of images in the archive. Hence automated analysis of the cervigram images by computerized tools is highly desirable. Automatically extracting visual features from different regions in the cervigrams, e.g. size, color, texture, shape and relative position within the cervix can be used for cervical cancer research, to assist in the training of experts and to enable future computerized cancer screening.

This study focuses on the extraction and segmentation of a specific tissue within the cervix, known as the acetowhite tissue (AW). The AW is a major indicator of cervical cancer.

A. Challenges to automated cervigram analysis

A typical cervigram is presented in Fig. 1. The *cervix region*, which is the main region of interest within the cervigram, is located in the central part of the image with the surrounding

Copyright (c) 2009 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org

A. Alush is with the Department of Biomedical Engineering, Tel-Aviv University, Tel Aviv 69978, Israel. E-mail: amiralus@post.tau.ac.il

H. Greenspan is with the Department of Biomedical Engineering, Tel-Aviv University, Tel Aviv 69978, Israel. E-mail: hayit@eng.tau.ac.il

J. Goldberger is with the School of Engineering, Bar-Ilan University, Ramat-Gan 52900, Israel. E-mail: goldbej@eng.biu.ac.il

vaginal walls and parts of the clinical equipment, such as the speculum or swab intruding in the image. A dark surrounding frame can be seen containing lines and text that are overlaid on the image at the time of the photographic development. The cervix region is defined by the *cervix boundary*. Automated detection of the cervix boundary defines the region of medical and anatomical interest within the cervigram and enables further analysis to focus on the cervix region itself.

The main tissues of interest as defined by NCI experts are: 1) The *Squamous Epithelium* (SE) - the normal cervix tissue. It appears as a homogenous pinkish-tan region and consists of multiple layers of cells (Sellors and Sankaranarayanan [37]); 2) The *Columnar Epithelium* (CE), which everts out of the Os (the opening of the cervix) when the cervix grows rapidly and enlarges under the influence of estrogen (after menarche and during pregnancy). This tissue is characterized by its bright red color and a rough textured appearance. The CE region is not always visible within a cervigram image; 3) The *acetowhite* (AW) region, the main focus of this work, is epithelium that turns white and is visible for a period of time following the application of acetic acid and is a major visual indicator of cervical cancer. Invasive squamous-cell cervical cancers are preceded by a long phase of preinvasive disease, collectively referred to as cervical intraepithelial neoplasia (CIN). CIN may be categorized into grades 1, 2 and 3 depending upon the proportion of the thickness of the epithelium showing mature and differentiated cells (Reid and Scalzi [34]). Areas of CIN in the AW region undergo maximal coagulation due to their higher content of nuclear protein and prevent light from passing through the epithelium. As a result, the subepithelial vessel pattern is obliterated and less easy to see and the epithelium appears white. This reaction produces a noticeable effect compared to the normal pinkish color of the surrounding normal squamous epithelium.

Several other regions may appear within the cervix image including mucus, blood stains and Squamous Metaplasia (SM). The SM refers to the physiological replacement of the everted columnar epithelium on the external cervix by a newly formed squamous epithelium. This healthy tissue may appear as white regions that can be confused with AW lesions. The region of the cervix where squamous metaplasia occurs is referred to as the transformation zone.

The automated extraction of the AW region is a very complex and challenging task due to a variety of factors: Several artifacts are generated during the acquisition process. Due to the powerful flash of the camera and the convex shape of the cervix the image tends to be brighter around the cervix center and the illumination decreases gradually towards the cervix boundary. This results in an inhomogeneous appearance within and across the tissues, which automatic segmentation algorithms fail to differentiate correctly. In particular, bright regions may be misclassified as AW lesions, whereas AW lesions located in the shaded regions are not detected. Additional artifacts that interfere with tissue segmentation are known as specular reflections (SR). These artifacts are seen as small and bright regions on the cervix surface, which are generated during the image acquisition process due to the presence of fluids (see Fig. 1). Another factor is that fact that the cervigram

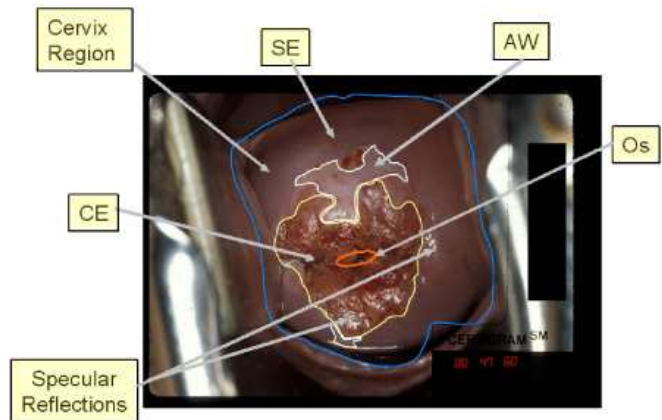


Fig. 1. A typical cervigram: marked are the cervix region, the columnar epithelium (CE), the squamous epithelium (SE), the acetowhite (AW), the entrance to the endocervical canal (Os) and the specular reflection artifacts.

archive varies considerably in that the image acquisition setup is not constant. The viewing angle varies significantly across images causing the cervix region to differ in intensity and shape from one image to another. In addition, the physical scene that is imaged has intrinsic variability. For example, in different patients the cervix is not the same size and additional non-cervical tissues or medical instruments may be present. One key difficulty is the variability of cervix tissue content within the images, as not all defined tissue types are present in each cervigram image. Fig. 2 shows examples of cervigrams that illustrate the content variability of the images within the archive. Note the variability in tissue color, texture, cervix size and relative position of the cervix in the image.

Due to the large variation in image appearance in the archive, features such as intensity, color and texture of a certain tissue type in one image often overlap with that of a different tissue type in another image. This makes reliable tissue segmentation in a large number of images a very challenging problem. In addition, intensity modeling within a single image is problematic: Although AW lesions tend to be brighter than their immediate surroundings they are not necessarily the brighter area in the image. As a result, it has been concluded that the features mentioned above, also known as region-based features, are not significant enough for the AW lesion segmentation task [12].

To overcome this problem our detection method is based on *boundary* cues. Relevant boundaries are learned in a supervised manner using a training set of labeled images where the object's boundaries are manually drawn. The ability to detect an object using only its boundaries can be highly advantageous when it is impossible to differentiate between object regions and non-object regions as in the case of cervigram images. Our approach is based on three components. The first is the watershed transform that converts the image into an edge map that contains the lesion boundary. The second is a feature-free boundary representation method that is based on visual words and a visual dictionary. The third is viewing the watershed map as a Markov Random Field (MRF) [20] in which each watershed superpixel corresponds to a binary random variable

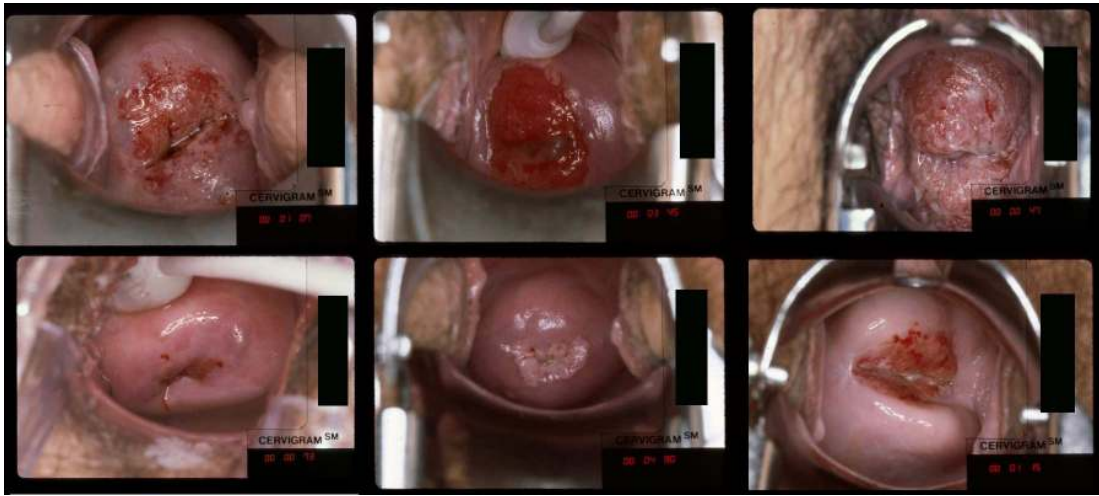


Fig. 2. Examples of cervigrams images in our dataset illustrating the existing variability among the AW regions.

indicating whether the superpixel is part of the lesion. The final segmentation is obtained by applying a belief-propagation algorithm on the loopy MRF. The methodology presented is fully automatic. If desired, user input can be incorporated in an interactive mode, with user cues incorporated within the MRF representation.

The organization of this article is as follows. Section II provides an overview of related work. Section III describes the parameter learning phase based on labeled images. Section IV presents an overview of the object detection/segmentation algorithm. The experimental validation is described in Section V. In Section VI the algorithm is extended to interactively incorporate user markers.

II. RELATED WORK

A. Cervigrams: related previous work

Previous work on the analysis of cervigram images has dealt primarily with automated landmark extraction, including the extraction of the cervix boundary, detection of the Os, and detection (and elimination) of specular reflections [43] [12]. The task of tissue segmentation, and in particular, the clinically important task of AW lesion detection and segmentation, remains, as yet, unsolved. Initial studies can be found on the analysis of individual cervigram images, or higher-resolution colposcopic images. Most of these studies are semi-automated, requiring the user to mark regions of interest on various cervix tissues (Cristoforoni *et al.* [5], Ji *et al.* [18], Pogue *et al.* [30]). Features such as color [30] texture [18] and shape [5] are then extracted. Based on these features the manually extracted regions are associated with different cervix tissues using various classifiers, such as neural networks [5] or the minimum distance classifier [18]. Other works have started to address the task of fully-automated colposcopic image analysis (Lange [23], Van-Raad *et al.* [32]). The data in these works were collected under controlled illumination conditions; thus variations in illumination are minimal.

Preliminary attempts at segmentation of the cervigrams in the NIH database were recently reported. Works to date have

mainly focused on one or two specific analysis tasks (e.g. specularity removal and region of interest detection [11], [41]). Initial results on AW segmentation, which is the focus of this work have been reported by a few groups: Srinivasan *et al.* [38] used texture features to represent vascular patterns inside the acetowhite regions. In [11] textured vs smooth regions in the image are separated by using a multi-scale texture-contrast feature per pixel, followed by clustering of the texture feature. Following the textured region separation, the smooth regions are modeled as a mixture of four Gaussians in the CIE-Lab color feature space. The cluster with the highest mean intensity is identified as a candidate AW lesion cluster. The methodology suggested is preliminary and is shown not to be robust to the large inherent variability in the data. Due to illumination effects the AW and SE tissues often possess very similar colors, and AW lesions are wrongly detected. Furthermore, AW lesions located in shaded areas of the image are not detected at all. In [16], SVM is used to perform color-based tissue classification in order to segment different tissue types, in particular focusing on the AW region. The segmentation performance is optimized with respect to the color space and granularity, with color spaces including RGB, HSV and CIE-Lab. Varying granularity in the representation includes individual pixel level colors, and colors of clustered regions (using mean-shift clustering). The work focuses on a comparison between the color space representations, and the granularity level, in terms of training time and memory usage, as well as defining an appropriate SVM kernel for the task. Qualitative segmentation results on a small image set are shown. The authors conclude that although SVM performs better than other classifiers used, the performance does not scale to large training sets. Large image sets are therefore not used and no quantitative segmentation results are provided.

The studies to date usually concentrate on one specific analysis task (e.g. single landmark or tissue) and report initial (qualitative) results with a small number of image examples. The large tissue overlap in feature space has hindered such attempts. Moreover, pixel based classification leads to numer-

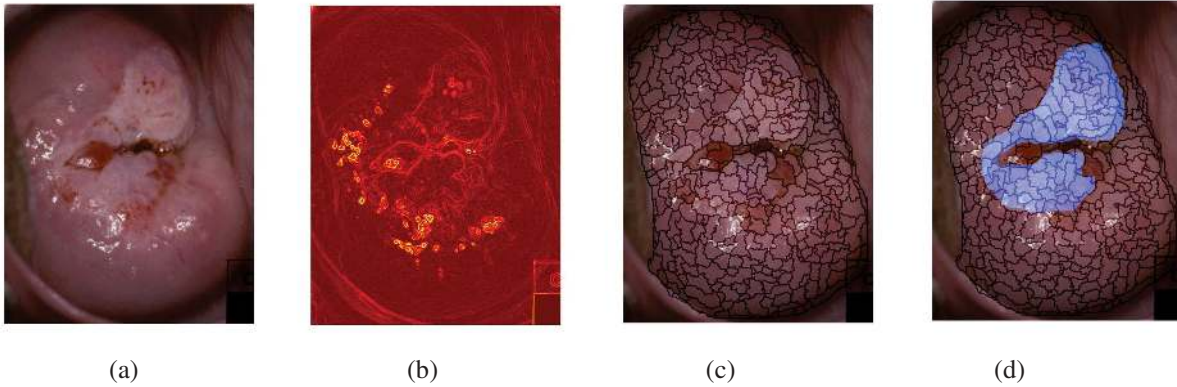


Fig. 3. Example of the watershed transform when applied to a cervigram image: (a) Original image; (b) Color gradients image; (c) Superpixels boundaries imposed on original image; (d) Expert's AW marking superimposed on superpixels map.

ous fragmented regions, many of which are false-positives. In the current study, we shift from region-based classification approaches to a boundary-based approach. The boundary is an important element in human expert lesion segmentation and can provide a great amount of unexplored information. In particular it can supply 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 studied here. It is also a common underlying difficulty in other challenging tasks involving tissue segmentation and lesion segmentation in general.

B. Object detection using boundary knowledge-related work

A substantial body of work in general scenery images uses boundaries for object detection [36], [17], [9], [2]. A boundary is a contour in the image plane which represents a change between separate surfaces or objects. Recent approaches [6], [26], [21] have developed a boundary learning paradigm applied to large natural image datasets in which the boundaries are manually annotated. Salient boundaries are classified against a background by combining local gradient cues, and a probabilistic edge map is produced. The approach described by [31], [1] is perhaps the most relevant to our study. It deals with the task of detecting object boundaries by first selecting candidate boundary pixels extracted by a standard edge detector (e.g. Canny [4]) and then applying the OBJ CUT segmentation algorithm [31], [22].

One promising approach to object segmentation is to use class specific object boundary learning with a higher level of knowledge of the image; i.e. an over-segmentation map. The watershed transform [40], a well established technique that partitions images into homogenous regions, can serve as a suitable platform for such a task [24], [10], [27]. A vast amount of work exists in the area of image segmentation and object detection using the watershed transform e.g. [13], [3], [10]. In the current work we use watershed segmentation for specific object boundary detection because it not only reduces the variability of negative examples, but can also lessen the

computational cost considerably.

III. LEARNING A BOUNDARY PROBABILISTIC MODEL

In this section we describe the pre-processing steps which translate an image into a probabilistic graphical model. We also describe the learning process of the parameters of the graphical model based on a training set of labeled images. We start the object boundary search from the watershed edge map. The next step is representing the edge pixels using patches centered at the pixel. In the last training step we use labeled data to build visual word histograms for edge pixels that correspond to the objects' contour, and for edge pixels in the complementary group. These pre-processing and training steps are described in detail below. Note that throughout this paper we refer to the acetowhite lesion as an "object", since we believe that the framework described here can also be adapted to the task of general object detection, even though this is beyond the scope of this work.

A. Extracting watershed edge maps

Each image is first over-segmented into superpixels by applying the watershed transform [40]. The watershed transform is a morphological segmentation tool that is applied to gray-scale images to solve a variety of image segmentation problems. The transform can be viewed as the flooding of a topographic surface (the input image), that extends from its regional minima while waters from different sources are prevented from merging. The final result is a set of segments or superpixels, separated by crest lines, or watershed arcs. The superpixels possess coherent region features and their boundaries are perpendicular to the image gradients. The underlying assumption is that the most important object boundaries are captured in the watershed edge map [24], [27]. This property of the watershed transform is crucial; otherwise attempts to achieve accurate object segmentation using its boundaries would be meaningless. The proposed framework thus considers only watershed edge pixels as the object boundary candidates. A classification scheme determines which of the candidate pixels are in fact located on the object boundary. The use of watershed edge pixels as candidates reduces

considerably both the variability of negative examples and the computational burden.

The watershed algorithm is illustrated in Fig. 3. Fig. 3(b) shows the normalized gradient map of the cervigram shown in Fig. 3(a). Fig. 3(c) presents the boundaries of the generated superpixels, imposed on the original cervigram. Note that the watershed edges do not cross any important tissues. In particular, there is an overlap with the expert-marked AW boundaries as shown in Fig. 3(d).

B. Building a visual word dictionary

The next step is feature extraction for each pixel in the watershed edge map. We apply a pixel-level variant of the “visual words” paradigm on the watershed edge pixels, to obtain a probabilistic model (histogram) for boundary and non-boundary pixels. The “visual words” paradigm is a recently introduced concept that has been successfully applied to scenery image classification tasks (see e.g. [8]). It is based on transforming the image into a set of visual words and representing the image (or objects in the image) using the statistics of the occurrence of each word as feature vectors. These visual words are image patches (small sub images) that are clustered to form a dictionary consisting of a small set of representative patches. This approach eliminates the need for explicitly specified boundary detection features. Instead, the features are implicitly found as part of the learning step composed of building the visual dictionary.

The process is done by extracting patches of size $n \times n$ for each pixel on the watershed edge map. The patches 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 [31]. 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 global gray-level differences.

We collect these edge patches from a set of training images. 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 clusters the data vectors in the projected space into K groups. Finally, the centroid of each group is taken to form a dictionary with K visual words. These visual words can be viewed as the average patches, hopefully representing all possible watershed edge patches in the training dataset. Note that this dictionary learning step is conducted in an unsupervised mode without any reference to the label (boundary/non-boundary) of each patch. A sample dictionary containing 100 visual words based on ten images is shown in Fig. 4. In the visual word approach, patches are assigned to the nearest cluster center, the residuals are discarded. Preserving all patches and using continuous measures is more information preserving than building a discrete dictionary. One of the interesting discoveries in this domain, as shown in large scale general object recognition, is that a discretized patch representation (binning) is sufficient [8], and the corresponding computational efficiency is unbeatable. Thus, in medical

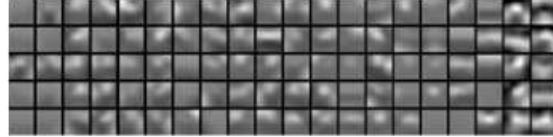


Fig. 4. A sample dictionary with 100 visual words representing all possible watershed edge patches, in the training dataset.

imaging in general, and in this paper, we follow the current state-of-the-art and utilize discrete distribution to model the patch behavior. The benefit of binning is the ability to utilize a discrete distribution (histogram) that can be easily trained and used. Without binning we have to deal with multimodal continuous distributions that are much more complicated for statistical learning and inference.

C. Building boundary histograms

Based on the labeled training set, the object’ 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 Euclidean 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 edge 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 each word from the dictionary is used in watershed edge pixels that are part of the object boundary and the second histogram is similarly defined for non-object boundary. 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. As a result of the training step, we obtain two probabilistic models for pixels on the watershed edge map. The first model describes the statistics of boundary pixels and the second one describes the statistics of non-boundary pixels. The training step is summarized in the following algorithm box.

Building Boundary and Non-Boundary Histograms:

- 1) Apply the watershed algorithm on the training images to obtain superpixels edge maps.
- 2) Extract a patch for each watershed edge pixel, rotate it and normalize it.
- 3) Apply the PCA algorithm on the patch vectors.
- 4) Apply the K-means clustering algorithm on the PCA projected space to obtain a visual dictionary and assign each edge pixel to a dictionary word.
- 5) Build a word frequency histogram for boundary pixels and for non-boundary pixels.

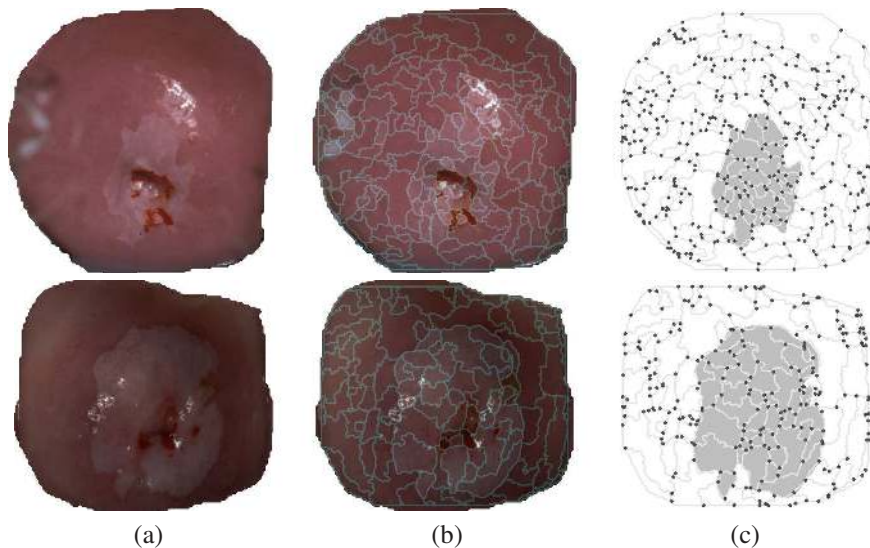


Fig. 5. MRF construction (2 examples): (a) Input image cropped to region of interest. (b) The watershed map overlaid on image. (c) The arc-level MRF. Ground truth lesion delineation shown in gray.

IV. THE LESION SEGMENTATION ALGORITHM

In the previous step we described the parameter learning phase which was based on a training set of labeled images. In this section we describe the object (lesion) detection algorithm when applied to a test image.

A. A probabilistic arc-level boundary in a new image

Given a new image, our goal is to detect and segment the object of interest. First, we utilize the watershed transform to divide the image into superpixels. Next, each one of the watershed edge pixels is translated into one of the visual words from the dictionary that was learned in the training step. This is done by rotating the patch vector centered at the edge pixel, normalizing 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 Euclidean distance).

Given an arbitrary edge pixel, the probability that this pixel was taken from an object boundary can be computed using Bayes' rule:

$$p(\text{boundary}|u) = \frac{p_{\text{boundary}}(u)}{p_{\text{boundary}}(u) + p_{\text{non-boundary}}(u)} \quad (1)$$

where u is a word index of the patch centered at the given pixel. Hence, we can compute the posterior probability for each patch to be on the object boundary. The result is a boundary/no-boundary separate decision for each watershed edge pixel.

However, one of our main claims is that being a part of a boundary is not an attribute of a single pixel. Assuming that the object boundary is a part of the watershed edge map, either all the pixels in a given watershed 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 pixel-level boundary probabilities into arc-level probabilities. We can, for example, define the arc probability given that it is a part of the object boundary to be the product of all the probabilities of the pixels on that arc. This approach, however, was found to be too aggressive and makes it almost impossible to overcome erroneous local boundary decisions. Instead, 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:

$$p(\text{arc}|\text{boundary}) = \frac{1}{|\text{arc}|} \sum_u p_{\text{boundary}}(u) \quad (2)$$

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

B. The arc-level watershed MRF

Next we use the watershed arc-level probability map for object segmentation. Our goal is to segment the image into object (lesion) and background. To accomplish this we classify each watershed superpixel either as a part of the object or as a part of the background. We use the arcs of the watershed edge map as indicators to whether two adjacent superpixels have the same object/background label. An arc which according to the extracted features, is more likely to be a boundary is an indication of neighboring superpixels with different labels. An arc which is more likely to be a non-boundary is an indication of neighboring superpixels with the same labels. To translate this intuition into a rigorous mathematical model we view the watershed superpixels as an MRF [20] such that each arc

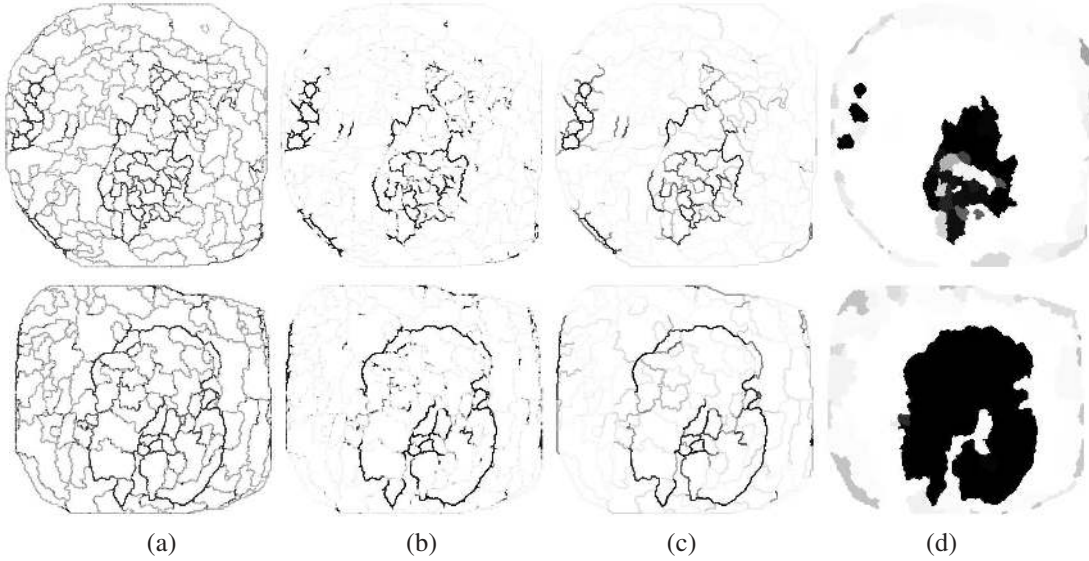


Fig. 6. Examples of the main algorithm steps (Corresponding original cervigram image is shown in Fig. 5): (a) Pixel-level AW/non-AW probabilities. (b) Results of the pixel-level MRF optimization. (c) Arc-level AW/non-AW probabilities; (d) Final segmentation results.

contributes a two-variable factor. We next provide a formal description of this MRF.

Let $x = \{x_1, \dots, x_n\}$ be a set of jointly distributed binary random variables associated with the watershed superpixels where n is the number of superpixels induced by the watershed transform. The assignment $x_i = 1$ means that the i -th superpixel is part of the object and $x_i = 0$ means that i -th superpixel is outside the object. Hence, each assignment of binary values to x_1, \dots, x_n corresponds to an object/background segmentation of the underlying image. From the previous step of the algorithm we obtain local information on each arc. For an arc between adjacent superpixel i and superpixel j we use the following notation for the corresponding factor:

$$\phi_{ij}(x_i, x_j) = \begin{cases} p(\text{arc}_{ij}|\text{no-boundary}) & \text{if } x_i = x_j \\ p(\text{arc}_{ij}|\text{boundary}) & \text{if } x_i \neq x_j \end{cases} \quad (3)$$

where arc_{ij} is the arc between the two superpixels and $p(\text{arc}_{ij}|\text{boundary})$ is defined in Eq. (2). The joint probability function is:

$$p(x_1, \dots, x_n) = \frac{1}{Z} \prod_{\{i,j\} \in E} \phi_{ij}(x_i, x_j) \quad (4)$$

where E is the set of all the arcs in the watershed edge map and Z is a normalization scalar. The preferred binary configuration is the 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. The optimal segmentation is the most likely configuration:

$$\hat{x} = \arg \max_{x \in \{0,1\}^n} p(x) \quad (5)$$

C. Arc-level gradient information

Even though the boundary information can separate the image into two distinct parts, it still lacks the ability to decide

which part is the foreground and which is the background. For each pair of superpixels i, j we have $\phi_{ij}(x_i, x_j) = \phi_{ij}(1 - x_i, 1 - x_j)$ and therefore for each possible object segmentation x_1, \dots, x_n it can be easily verified that $p(x_1, \dots, x_n) = p(1 - x_1, \dots, 1 - x_n)$. Hence, there is a global degree of freedom in the model that affects the meaning of the binary values. This degree of freedom can cause a severe problem when it turns into a local one where in some image areas '1' corresponds to the object and in other areas '1' corresponds to the background. To overcome this difficulty more knowledge must be introduced. We observed that most gradients along the acetowhite boundaries point inward, since for the most part the acetowhite, as its name indicates is brighter than its immediate surroundings. Note that it is not true that globally the lesions are the brightest area in the image; this is only a local feature. Using the cervigram labeled train set, we compute the mean and variance of the gradients in a local environment of 3×3 around each watershed pixel. To get arc-level information we average all the gradients associated with the pixels on a given arc of the watershed edge map. We model the watershed map arc-level gradient information using three normal distributions depending on whether the gradient is inside the object, on the boundary or outside the object:

$$\begin{aligned} \text{gradient inside object} &\sim \mathcal{N}(0, \sigma_{11}^2) & (6) \\ \text{gradient on the background} &\sim \mathcal{N}(0, \sigma_{00}^2) \\ \text{gradient from object to background} &\sim \mathcal{N}(\mu_{01}, \sigma_{01}^2) \end{aligned}$$

The model parameters are learned using the labeled training images. Note that the average of the gradients inside the object is zero since we use gradients in both directions. The same argument applies for gradients outside the object. We show in the experiment section that only using gradient information yields poor results. However, combined with the visual-word

boundary information it breaks the object/background symmetry of the arc-level model.

We now return to the problem of segmenting a test image. For each arc of the watershed edge map we compute a gradient feature in the same way as described above. We incorporate the gradient information into the global probabilistic model via an additional pairwise factor. For each arc of the watershed map,

$$\psi_{ij}(x_i, x_j) = \begin{cases} \mathcal{N}(\text{grad}_{ij}; 0, \sigma_{00}^2) & \text{if } x_i = 0, x_j = 0 \\ \mathcal{N}(\text{grad}_{ij}; \mu_{01}, \sigma_{01}^2) & \text{if } x_i = 0, x_j = 1 \\ \mathcal{N}(\text{grad}_{ij}; -\mu_{01}, \sigma_{01}^2) & \text{if } x_i = 1, x_j = 0 \\ \mathcal{N}(\text{grad}_{ij}; 0, \sigma_{11}^2) & \text{if } x_i = 1, x_j = 1 \end{cases} \quad (7)$$

where grad_{ij} is the value of the gradient from superpixel i to superpixel j . Note that unlike the boundary factor $\phi(x_i, x_j)$, the gradient factor is not symmetric; i.e. $\psi_{ij}(x_i, x_j) \neq \psi_{ij}(1-x_i, 1-x_j)$. This removes the object/background symmetry of the arc-level model. Therefore, the optimal segmentation is the most likely configuration of the following MRF:

$$\prod_{\{i,j\} \in E} (\phi_{ij}(x_i, x_j) \psi_{ij}(x_i, x_j)) \quad (8)$$

D. The BP optimization

The marginal probabilities of the MRF variables (superpixels) correspond to the posterior probabilities to be a part of the object. Since the MRF graph has loops, it is not feasible to compute the exact marginal. The problem is known to be NP hard. Instead, we utilize the belief propagation (BP) approximation algorithm [29] which is an efficient way to solve inference problems in graphical models. The BP algorithm propagates information through the MRF via a series of messages sent between neighboring superpixels. There are two variants of BP, the sum-product and the max-product. In a loop-free MRF graph the max-product message update rule finds the most likely pattern (which corresponds here to the optimal segmentation). In our case of loopy MRF the max-product is an approximation algorithm and we can view the sum-product as a soft version of the max-product. The max-product message from superpixel i to a neighbour superpixel j is:

$$m_{i \rightarrow j}(x_j) = \max_{x_i} \phi_{ij}(x_i, x_j) \psi_{ij}(x_i, x_j) \prod_{k \in N(i) \setminus j} m_{k \rightarrow i}(x_i) \quad (9)$$

where $N(i)$ is the set of all the superpixels that share a common arc with superpixel i . In the sum-product variant of the BP algorithm the maximization in Eq. (9) is replaced by a summation. In the experiment section we show that for our task the sum-product segmentation performance is slightly better. The approximate marginal distribution of x_i , the belief of x_i , is:

$$\text{belief}_i(x_i) = \prod_{k \in N(i)} m_{k \rightarrow i}(x_i) \quad (10)$$

We iterate until the messages converge or until a predefined number of iterations is reached. As a final step we threshold the belief to obtain a hard-decision label for each superpixel.

The variables correspond to the superpixels and each arc contributes a factor.

The algorithm is robust to the message initialization method, thus a random initialization of all messages is sufficient. As pointed out earlier, erroneous boundary fragments are detected in the proximity of the cervix outer boundaries. This is caused by the presence of the vaginal walls which were not excluded in the expert' region of interest selection. The vaginal walls boundaries resemble the AW lesion boundaries. In order to avoid this artifact all messages from outside the region of interest are initialized as "background" with high probability.

Fig. 5 shows the watershed map and the manual expert segmentation of two image examples. Fig. 6 provides a visual demonstration of the steps of the automatic segmentation algorithm for these two examples. The proposed algorithm is summarized in the following algorithm box.

The Object Segmentation Algorithm:

- 1) Apply the watershed algorithm to obtain the superpixels and the edge map.
- 2) Compute a patch for each watershed edge pixel and assign it to a dictionary word.
- 3) Compute the probability of each watershed arc as a boundary element and as a non-boundary element.
- 4) Perform arc-level belief propagation and threshold the result to obtain an object segmentation:

- Initialize messages and iterate until convergence:

$$m_{i \rightarrow j}(x_j) = \sum_{x_i} \phi_{ij}(x_i, x_j) \psi_{ij}(x_i, x_j) \times \prod_{k \in N(i) \setminus j} m_{k \rightarrow i}(x_i)$$

- Compute the beliefs:

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

- Final segmentation:

$$x_i \in \begin{cases} \text{object} & \text{belief}_i(x_i) \geq 0.5 \\ \text{background} & \text{otherwise} \end{cases}$$

V. EXPERIMENTAL VALIDATION

The implementation of our method to the lesion detection task starts by applying the watershed transform. The pixels on the watershed edge map in the training images were taken as centers of patches of size 11×11 each. Applying PCA, the dimensionality was reduced to 10. The K-means algorithm was utilized to cluster the patches into a dictionary of $K = 100$ words. We found that these parameters work well in terms of both computational complexity and performance (see Section V-C for a discussion on the algorithm's parameter sensitivity). Next, using the AW and non-AW boundary pixels that were defined on the training images, the frequencies of every word in the dictionary for AW and non-AW boundaries were computed. In the training step, positive examples were

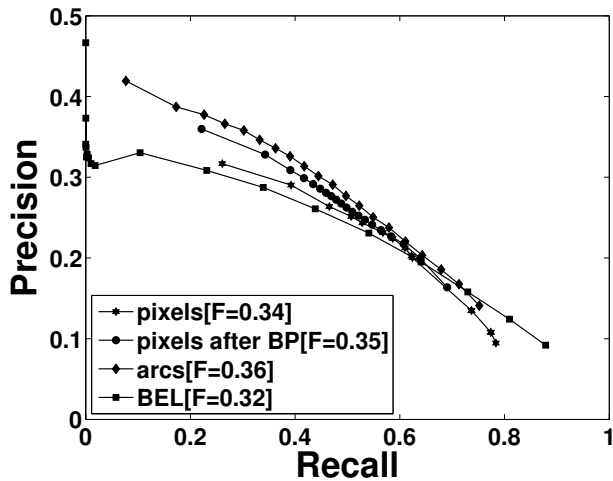


Fig. 7. Precision vs. Recall curves for performance evaluation; comparing variants of our approach to edge learning with “Boosted Edge Learning” [6].

considered to be the watershed pixels within a distance of at most three pixels from the ground truth annotation. Negative examples were the watershed edges that were at least 15 pixels away from the ground truth annotation. Pixels that were within these two thresholds were not used for the histograms learning process. To obtain a decision map for a new image, the same pre-processing as on the training images was applied. Each image then went through four major algorithmic steps.

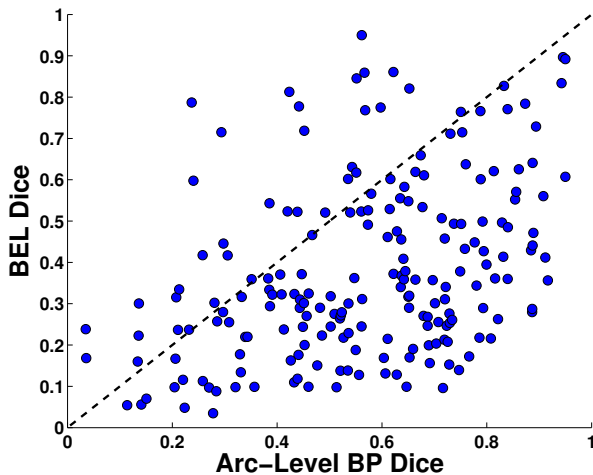


Fig. 9. A scatter plot of Dice BEL versus Dice arc-level BP.

The first step was to assign a visual word from the dictionary to each patch on the watershed edge map, which yielded the first step probabilistic map. Next, pixel-level MRF optimization was applied to obtain a smooth local map. Then, the pixel-level information was converted into arc level probabilities. Finally, in the fourth step the belief propagation algorithm was applied to the watershed arc-graph. Since the object segmentation was carried out 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.

The annotated data set we used contains 211 cervigram

images. In all the experiments described below, a single set of 11 images was used for training the system and finding its parameters. The other 200 images were used for testing. The gold standard for the automated segmentation results is expert manual segmentation.

The discussion of the algorithm’s performance is divided into two parts: the first deals with the supervised boundary extraction which yields in the arc level probabilities. The second, and more important part, is the final object (lesion) segmentation based on the arc-level belief propagation, which produces the superpixel classification. We begin our evaluation by examining the boundary extraction performance followed by an examination of the object segmentation performance.

A. Evaluating boundary extraction

A standard method for evaluating boundary extraction performance is the precision vs. recall framework described in [26].

We used a set of 20 threshold values. For each threshold the boundary points above the threshold were matched to the ground truth experts’ annotations and the precision vs. recall results were recorded. We compared the boundary extraction performance to the “Boosted Edge Learning” (BEL) [6] which is a classifier built on top of a probabilistic boosting tree (PBT) that uses approximately 5000 features across different scales, including canny edge results, Haar filter responses, gradients and curvature. We explored different parameters settings and found that the default parameters suggested by the authors provide the best results.

Fig. 7 shows precision-recall curves for boundary detection using the proposed approach and the BEL approach. Results using pixel-only and arc-based probabilities are included in the plot and demonstrate the improvement in the algorithm’s performance in each of the proposed algorithm steps. The results as shown in Fig. 7 demonstrate that for the edge detection task the proposed methodology is comparable to the state-of-the-art BEL approach. We also tried to smooth the pixels probabilities before averaging them, by applying a BP along the pixels of watershed edge map in such a way that adjacent pixels were encouraged to have the same boundary/non boundary label. We show in Fig. 7 that this operation slightly improves the final results.

B. Evaluating object (lesion) segmentation

We next evaluated the lesion region segmentation results of the proposed algorithm. The overlap segmentation measures between two possible segmentations are described next.

Let S be the area of the automatically segmented region, R the expert segmentation and \hat{R} its complement (the area outside the expert marked boundary). The following overlap measures were used for evaluation of the segmentation quality of a single object/region: the Dice measure: $2 \frac{|S \cap R|}{|S| + |R|}$, the Sensitivity measure: $\frac{|S \cap R|}{|R|}$ and the False Positives (FP) measure: $\frac{|S \cap \hat{R}|}{|\hat{R}|}$. We also used the Rand index [33], which is a standard non-parametric measure of clustering quality. The Rand Index

lies between 0 and 1. When the two partitions agree perfectly the Rand index is 1.

In the spirit of [42] the comparison was with a region-based classification variant of the BEL algorithm, which was trained using pixels on/off the object as true/false examples. The BEL was trained and tested using the exact same dataset we used to train/test our algorithm. Different BEL parameters settings were tested, including different patch size, tree depth and color variants, and all yielded similar results. The results we describe here are with the BEL default parameters. The automatic segmentation results of both our algorithm and BEL are summarized in Table I. Fig. 9 shows a scatter plot of comparative segmentation results of our method and the BEL algorithm for all the 200 images in the test set. Several examples of our segmentation results, the BEL results and the expert's annotations are shown in Fig. 8.

C. Parameters sensitivity analysis

We tested the algorithm's sensitivity to the following set of parameters (the selected values for each appear in brackets):

- 1) Number of dictionary words ($K = 100$).
- 2) PCA dimension ($d = 10$).
- 3) Patch size ($N = 11$).
- 4) Training set size ($T = 11$).
- 5) Patch preprocessing method (mean subtraction, patch rotation).
- 6) With/without pixel-level belief propagation (with).
- 7) Arc-level belief propagation using sum-product or max-product (sum-product).

Experiments were run per parameter. In the case of the training set size, the experiment was repeated 10 times with a random selection of the training images per run. Fig. 10 presents the results for the PCA size, dictionary size, patch size and training set size parameters. Table II presents the results for patch preprocessing methods and Table III shows sensitivity for arc-level BP variations. The main conclusions are as follows:

- 1) Patch size: when using a patch size larger than 7×7 similar results were obtained (sensitivity 0.55), while using a smaller patch size yielded slightly inferior results (sensitivity 0.52). If the patch size is too small (e.g. 5×5) not enough information surrounding the boundary is captured. Using larger patches increases the computational complexity without notable performance improvement. If the patch size is too large this can even impair performance since each patch is more unique and less similar to other patches in the image. This can be viewed as building a dictionary based on sentences rather than words.
- 2) Dictionary size: using 10 visual words yields a sensitivity of 0.48, while using more than 60 words a similar sensitivity of 0.55 was obtained. Using a small amount of visual words (e.g. 10 words), prevents the data from being accurately represented, while using more words does not significantly enhance the results but increases the computational complexity.

TABLE I
SEGMENTATION RESULTS OF THE ARC-LEVEL BP ALGORITHM AND BEL ALGORITHM.

	Dice	Sensitivity	Rand	False-Positive
arc-level BP	0.57	0.60	0.83	0.04
BEL(region)	0.37	0.23	0.65	0.21

TABLE II
ALGORITHM SENSITIVITY FOR THE PATCH PREPROCESSING METHOD.

Mean Subtraction	Patch Rotation	RAND	Dice	Sensitivity	FP
No	No	0.8	0.49	0.49	0.1
No	Yes	0.69	0.37	0.5	0.1
Yes	No	0.82	0.5	0.5	0.04
Yes	Yes	0.83	0.57	0.6	0.04

- 3) PCA size: Using fewer than 3 dimensions after the PCA reduction yields a dramatically inferior sensitivity of 0.1. When considering more dimensions a plateau (0.55) is reached. Using fewer than 3 dimensions is not sufficient and clearly the classifier cannot distinguish between background and foreground boundaries. Increasing the number of dimensions yields no noticeable improvement.
- 4) Training Size: Using an overly small train size (e.g. 1 training image) shows a sensitivity of 0.4, but using more train images clearly results in improvement. The algorithm shows robustness to train size, and no improvement is noticeable when further increasing the amount of training images. This is probably due to the nature of the data which present considerable variability even within the same image. Hence introducing more images does not lead to significant enhancement.

VI. INCORPORATING USER MARKERS

A. An interactive segmentation algorithm

So far we have 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. 11, in some cases the automatic segmentation is either lacking or erroneous especially in parts of the image that resemble the AW; this is one of the main challenges in this task. In such cases user knowledge can be utilized to improve the segmentation by incorporating user markers for the object and/or background. Using a simple user interface, see Fig. 12, 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 (hidden to the user) of the 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

TABLE III
ALGORITHM SENSITIVITY FOR ARC-LEVEL BP VARIATIONS

arc-level bp	RAND	Dice	Sensitivity	FP
sum product	0.83	0.57	0.60	0.04
max product	0.83	0.56	0.58	0.04

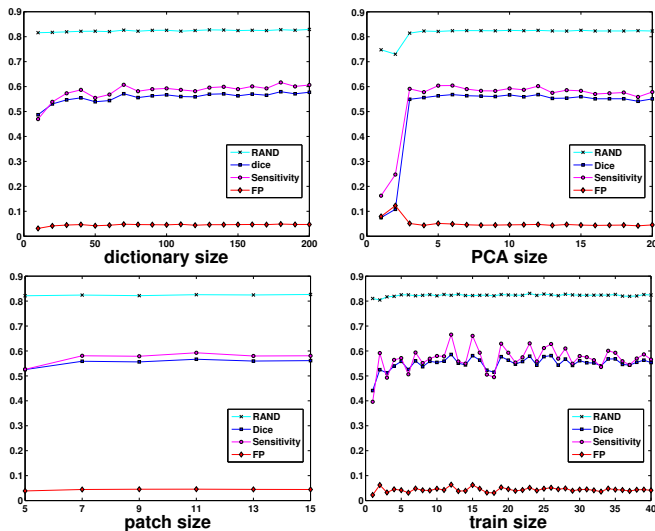


Fig. 10. Sensitivity analysis of the proposed algorithm’s to different changes in the parameter settings. A set of parameters ($K = 100, d = 10, N = 11, T = 11$) is kept fixed while on each parameter examination, only the relevant parameter varied and the segmentation performance on the test set is evaluated.

following form:

$$m_{i \rightarrow j}(x_j) = \phi_{ij}(x_i, x_j)\psi_{ij}(x_i, x_j) \quad (11)$$

where x_i is the object/background information (markers) provided by the user on superpixel i . All other messages related to superpixels that were not marked by the user remain as defined in Eq. (9). Note that the final ‘belief’ of superpixels marked by the user is set to be the value specified by the user instead of using Eq. (10).

If the user is still not satisfied with the quality of the segmentation he/she can input additional markers. From our experience it usually takes at most two interactive iterations to obtain the desired segmentation. Once the user provides (objcte/background) markers it takes the system just few milliseconds to compute the updated segmentation.

B. Evaluating the algorithm’s interactivity

To test whether the interactive segmentation algorithm improves the automated segmentation when correct markers are entered, we ran a quantitative evaluation on our entire test set using a simulation that automatically found all discrepancies between the expert’s segmentation and our automated segmentation for both background and foreground. It then randomly chose an increasing number of background/foreground seed points from each image. The results for each step were then recorded. An illustration of the disagreement superpixels and the appropriate background/foreground seed points is presented in Fig. 13. An example of the simulated interactive segmentation process is shown in Fig. 14. Note how the final interactive segmentation result in 14(f) closely resembles the expert’s segmentation shown in Fig. 13(b). Note that not all available seed points (shown in Fig. 13) are used to obtain the final segmentation result.

A statistical summary of the interactive segmentation simulation results when applied to our entire test set is presented in

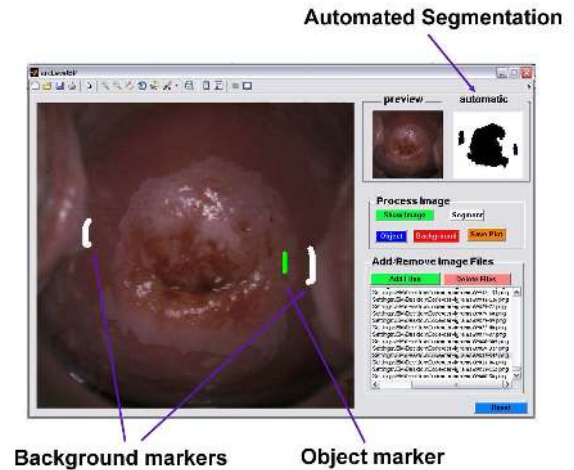


Fig. 12. The user interface for interactive segmentation. A preview of the automated segmentation is shown in the top right corner; the user is then able to set object markers (green) and/or background markers (white). The markers can be either mouse strokes as shown in the figure, or mouse clicks.

Table IV. We used the same parameter set in the beginning of this section. It is clear from the results that the interactive segmentation further improves the algorithm’s performance with each click introduced. We also show several examples of interactive segmentation performed manually using mouse strokes in Fig. 11.

TABLE IV
INTERACTIVE SIMULATION RESULTS. THE MEAN RESULTS FOR OUR ENTIRE TEST SET, USING AN INCREASING NUMBER OF RANDOMLY CHOSEN OBJECT/BACKGROUND SEED POINTS (MARKERS).

# seeds	Dice	Sensitivity	Rand	False-Positive
0 (automatic)	0.57	0.60	0.83	0.04
1	0.58	0.66	0.84	0.05
2	0.60	0.68	0.85	0.05
3	0.62	0.70	0.86	0.06
4	0.63	0.72	0.86	0.06
5	0.65	0.73	0.86	0.06

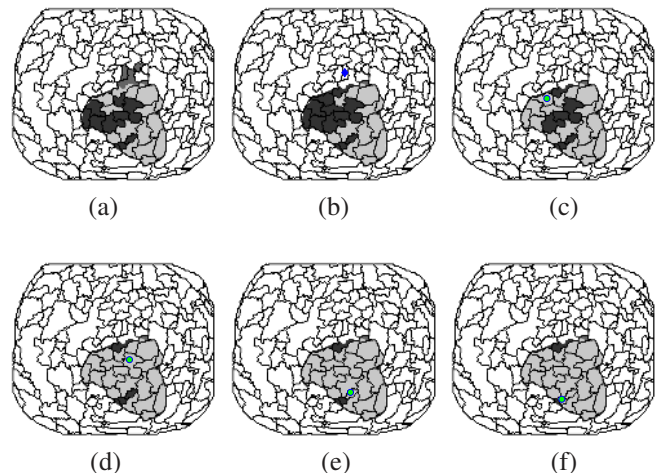


Fig. 14. The progression of the simulated interactive segmentation process. (a) is the automated segmentation and (f) is the final interactive segmentation after using 5 seed points.

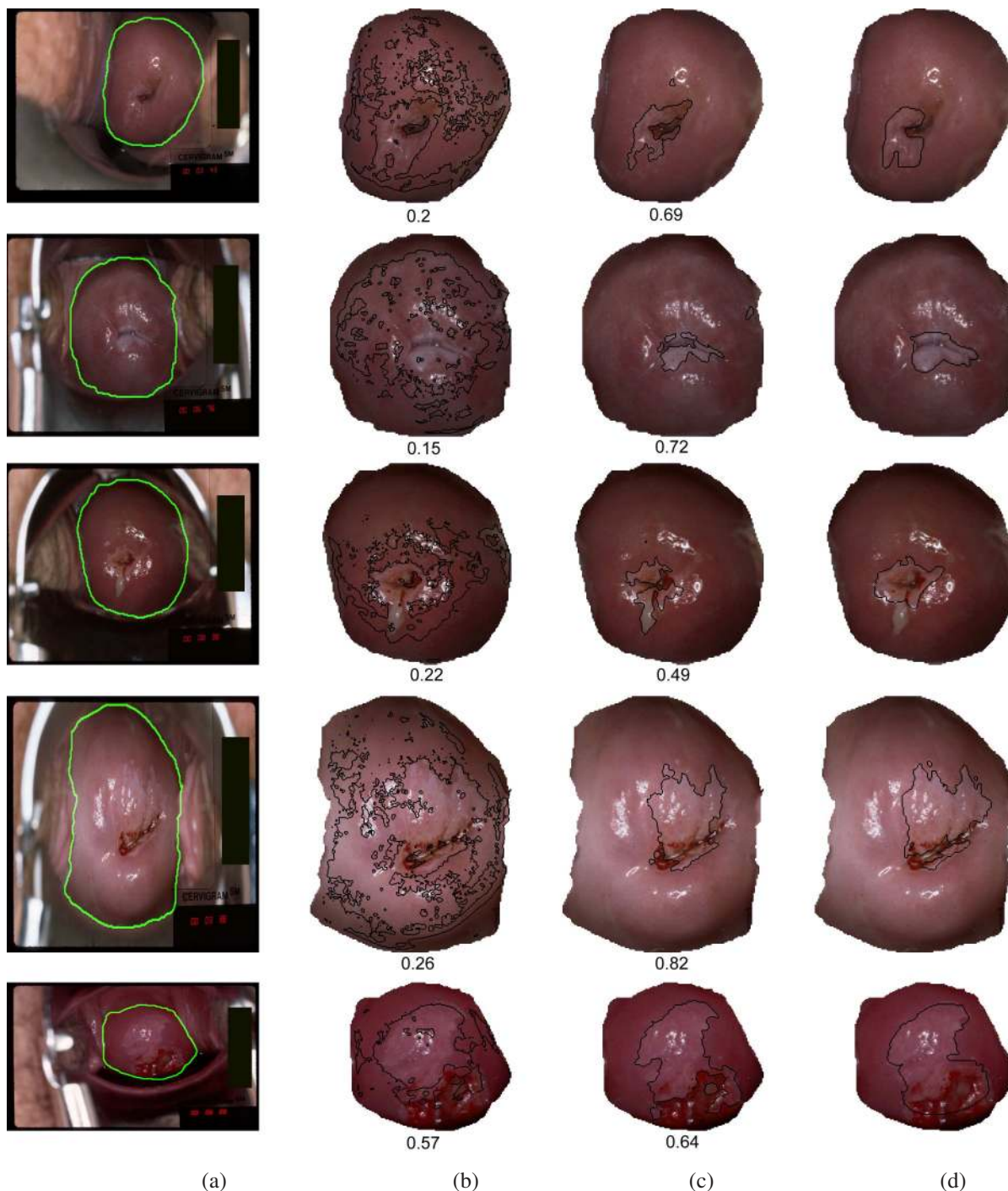


Fig. 8. Automatic segmentation results, a threshold of 0.5 is used for both algorithms. Segmentations delineated in black: (a) Input image. Region of interest marked in green. (b) BEL segmentation. (c) Our segmentation. (d) Expert's segmentation. The Dice performance measure is written below the segmentation results in (b) and (c).

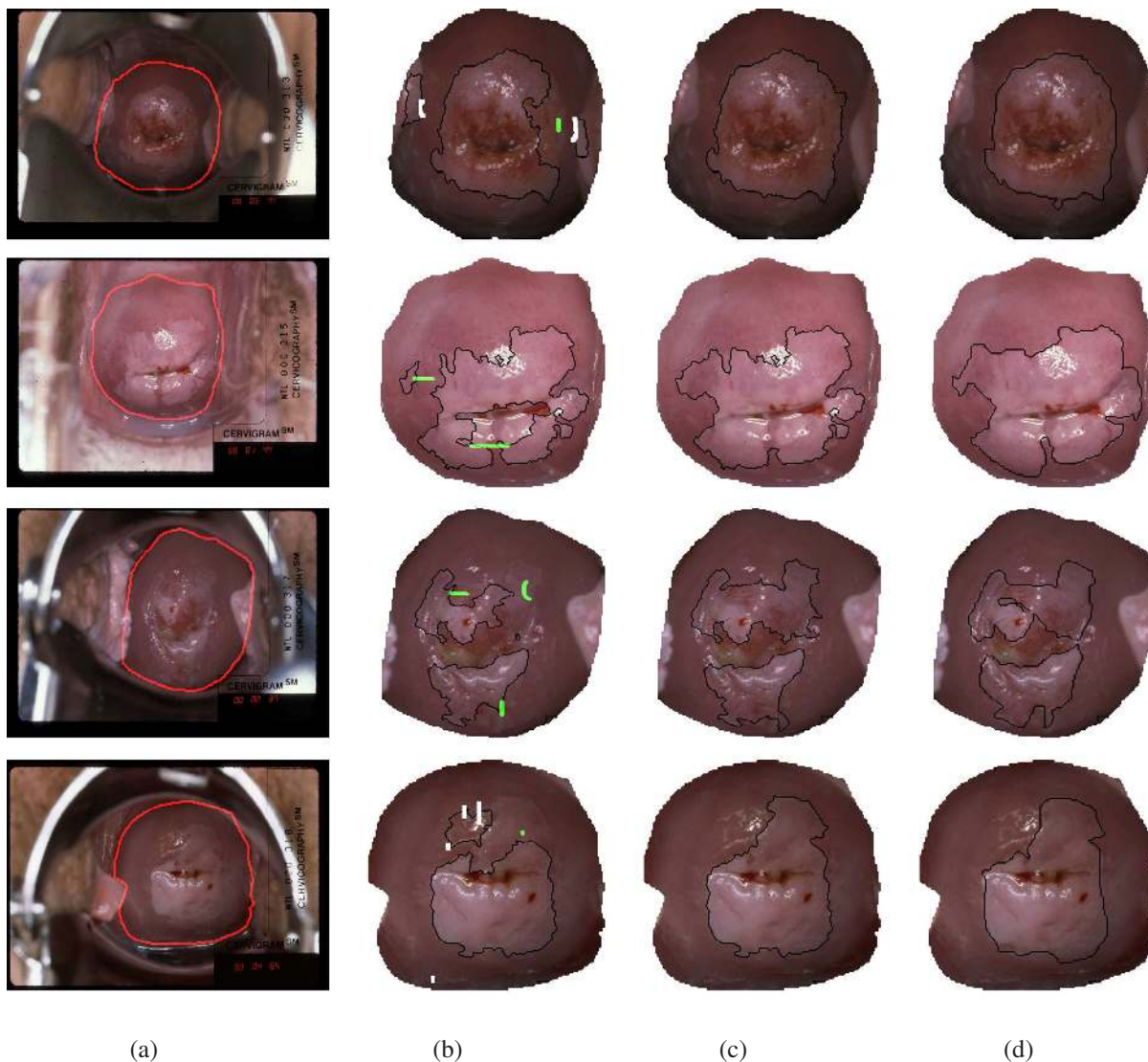


Fig. 11. Interactive segmentation results: (a) Input image. Region of interest marked in red. (b) Automatic segmentation (black line). User markers are colored: green (object), white (background). (c) The final segmentation result (black line). (d) Expert segmentation (black line).

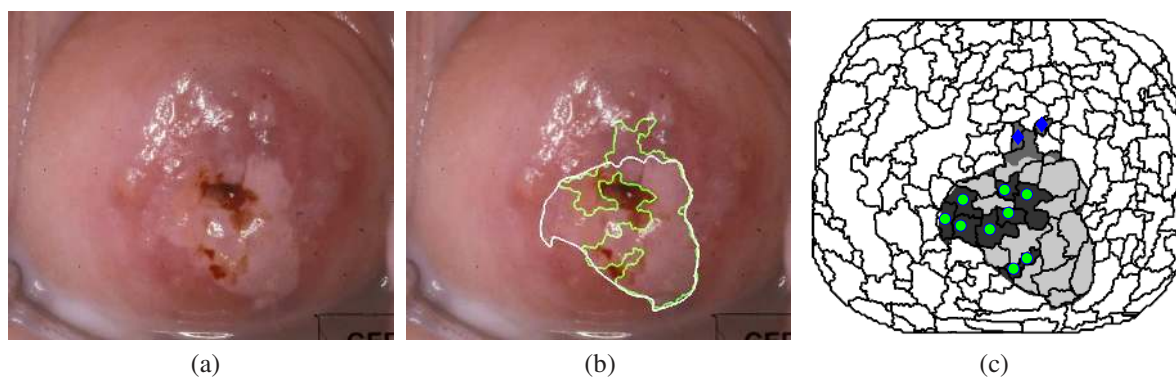


Fig. 13. The discrepant superpixels between the expert's segmentation and our automated segmentation. (a) Input image. (b) In white the expert's segmentation and in green our automated segmentation. (c) Both segmentations imposed on one another. The green circles represent a simulated seed points for the foreground and the blue diagonals represent seed points for the background.

VII. DISCUSSION

In this study we presented an automatic extraction and segmentation methodology (with an interactive extension) for lesion regions in uterine cervix images. The cervigram segmentation challenge has received attention from several research teams in the last decade. Publications to date show substantial results on various landmarks within the cervigram image, with a few qualitative AW segmentations, that are significantly discrepant from the expert ground truth. Segmenting out the lesion region remains the main motivation and challenge with minimal results published up to now. 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 results in the current work constitute a major step forward in automated segmentation capabilities and are a direct result of incorporating the boundary cues within the segmentation process, in particular in cases in which the region information is not sufficient. Based on the data-set we have checked, we can say that the automatic segmentation is usually very similar to the expert marking. In case the user is not satisfied with the automatic results, the interactive phase of our algorithm can be utilized to obtain the desired segmentation. The interactive phase is very fast and user friendly. Overall, based on our experience with the system, we can say that the AW tissue is extracted with a high degree of accuracy (as compared to a human expert). As such, it is now possible to automatically detect the presence of AW regions and to quantify their size. The AW tissue is a major indicator of cervical cancer. With the ability to extract the AW region, various screening tools can be developed for screening of the disease. Moreover, the capabilities shown in the work can facilitate access to the database of digitized cervix images which, as mentioned in the Introduction, is seen as a critical component by NCI and NIH, in training and in cervical cancer research.

Several issues can be explored further: 1) In the current work, boundary information is used with no added region-based input. In cases in which region-based information is available and of value, it can easily be incorporated into the MRF as additional single-variable factors. 2) Currently, in the building of the visual word dictionary (Section III-B), the features are made rotationally invariant up to a flip factor. A question arises as to the possibility of making the patch feature completely rotation-invariant. This choice is currently not taken due to the fact that in the feature extraction step we do not want to make any hard decision regarding the boundary orientation. Moreover, most of the patches in the watershed edge map are not on the lesion boundary. For those patches, inside/outside direction is meaningless. At the preprocessing feature extraction step we don't know if the patch is on the boundary, hence we cannot flip it to the 'correct' side at this step. An interesting alternative is to use the flip factor information at the MRF constructing step by rotating the patch such that the object side will be always, for example, on the left. We can then build a frequency histogram based on the training data. This alternative will be investigated at a future version of the system. 3) In the current study the training was

based on a single expert boundary marking for each training image. A generalizing to the case of multiple expert markings is of interest.

The method presented is a general one. We are currently testing its application to additional complex medical image segmentation tasks such as lesion segmentation in the liver and brain.

REFERENCES

- [1] O. Barinova, A. Kuzmishkina, A. Vezhnvets, and V. Vezhnvets. Learning class specific edges for vanishing point estimation. *Proc. of Graphicon*, pages 162–165, 2007. 4
- [2] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24:509–522, 2002. 4
- [3] Y. Boykov and G.F-Lea. Graph cuts and efficient N-D image segmentation. *Int. J. Comput. Vision*, 70(2):109–131, 2006. 4
- [4] J. Canny. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8:679–698, 1986. 4
- [5] P. M. Cristoforoni, D. Gerbaldo, A. Perino, R. Piccoli, F. J. Montz, and G. L. Captiano. Computerized colposcopy: Results of a pilot study and analysis of its clinical relevance. *Obstet. Gynecol.*, 85:1011–1016, 1995. 3
- [6] P. Dollar, Z. Tu, and S. Belongie. Supervised learning of edges and object boundaries. *Computer Vision and Pattern Recognition (CVPR)*, 2006. 4, 9
- [7] J. Eluf-Neto and C.M. Nascimento. Cervical cancer in Latin America. *Seminars in Oncology*, 28(2):188–97, 2001. 1
- [8] L. Fei-Fei and P. Perona. A bayesian hierarchical model for learning natural scene categories. *Computer Vision and Pattern Recognition (CVPR)*, pages 524–531, 2005. 5
- [9] P. Felzenszwalb. Learning models for object recognition. *Computer Vision and Pattern Recognition (CVPR)*, pages 56–62, 2001. 4
- [10] T. Geraud, J.-F. Mangin, I. Bloch, and H. Maitre. Segmenting internal structures in 3D MR images of the brain by markovian relaxation on a watershed based adjacency graph. *Int. Conference on Image Processing*, 1995. 4
- [11] S. Gordon, G. Zimmerman, R. Long, S. Antani, J. Jeronimo, and H. Greenspan. Content analysis of uterine cervix images: Initial steps towards content based indexing and retrieval of cervigrams. In *Proc. of SPIE Medical Imaging*, volume 6144, pages 1549–1556, 2006. 3
- [12] 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. 2, 3
- [13] K. Haris, S. N. Efstratiadis, N. Maglaveras, and A. K. Katsaggelos. Hybrid image segmentation using watersheds and fast region merging. *IEEE Trans. on Image Processing*, 7(12):1684–1699, 1998. 4
- [14] R. Herrero, A. Hildesheim, C. Bratti, M. E. Sherman, M. Hutchinson, J. Morales, I. Balmaceda, M. D. Greenberg, M. Alfaro, R. D. Burk, S. Wacholder, M. Plummer, and M. Schiffman. Population-based study of human papillomavirus infection and cervical neoplasia in rural Costa Rica. *Journal of the National Cancer Institute*, 92(6):464–474, 2000. 1
- [15] R. Herrero, M. H. Schiffman, C. Bratti, A. Hildesheim, I. Balmaceda, M. E. Sherman, M. Greenberg, F. Cardenas, V. Gomez, K. Helgesen, J. Morales, M. Hutchinson, L. Mango, M. Alfaro, N. W. Potischman, S. Wacholder, C. Swanson, and L. A. Briton. Design and methods of a population-based natural history study of cervical neoplasia in a rural province of Costa Rica: the Guanacaste Project. *Pan American Journal of Public Health*, 1(15):362–375, 1997. 1
- [16] X. Huang, W. Wang, Z. Xue, S. Antani, L.R. Long, and J. Jeronimo. Tissue classification using cluster features for lesion detection in digital cervigrams. In *Proc. of SPIE Medical Imaging*, 2008. 3
- [17] D. Huttenlocher, G. Kl, and W. Rucklidge. Comparing images using the Hausdorff distance. *IEEE Trans. Pattern Anal. Mach. Intell.*, 15:850–863, 1993. 4
- [18] Q. Ji, J. Engel, and E. Craine. Texture analysis for classification of cervix lesions. *IEEE Transaction on Medical Imaging*, 19(11):1144–1149, 2000. 3
- [19] T.C. Wright Jr. Cervical cancer screening using visualization techniques. *Journal of the National Cancer Institute Monographs*, 31:66–71, 2003. 1

- [20] R. Kindermann and J. L. Snell. *Markov Random Fields and their applications*. Providence, Rhode Island: American Mathematical Society, 1980. 2, 6
- [21] S. M. Konishi, A. L. Yuille, J. M. Coughlan, and S. C. Zhu. Fundamental bounds on edge detection: An information theoretic evaluation of different edge cues. In *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*, pages 573–579, 1999. 4
- [22] M.P. Kumar, P.H.S. Torr, and A. Zisserman. OBJ CUT. *Computer Vision and Pattern Recognition (CVPR)*, pages 18–25, 2005. 4
- [23] H. Lange. Automatic detection of multi-level acetowhite regions in RGB color images of the uterine cervix. In *Proc. of SPIE Medical Imaging*, volume 5747, pages 1004–1017, 2005. 3
- [24] Y. Li, J. Sun, C.-K Tang, and H.Y. Shum. Lazy snapping. *ACM SIGGRAPH*, pages 303–308, 2004. 4
- [25] L. R. Long, S. Antani, J. Jeronimo, M. Schiffman, M. Bopf, L. Neve, C. Cornwell, S. R. Budhihas, and G. R. Thoma. Technology for medical education, research, and disease screening by exploitation of biomarker in a large collection of uterine cervix images. In *Proc. of the 19th IEEE Symposium on Computer-Based Medical Systems (CBMS)*, 2006. 1
- [26] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. *Int'l Conf. Computer Vision*, pages 416–423, 2001. 4, 9
- [27] J. Park and J.M. Keller. Snakes on the watershed. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(10):1201–1205, 2001. 4
- [28] D.M. Parkin, F.I. Bray, and S.S. Devesa. Cancer burden in the year 2000. the global picture. *European Journal of Cancer*, 37:4–66, 2001. 1
- [29] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco: Morgan Kaufmann, 1988. 8
- [30] B. W. Pogue, M. A. Mycek, and D. Harper. Image analysis for discrimination of cervical neoplasia. *Journal of Biomedical Optics*, 5(1):72–82, 2000. 3
- [31] 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. 4, 5
- [32] V. Van Raad, Z. Xue, and H. Lange. Lesion margin analysis for automated classification of cervical cancer lesions. In *Proc. of SPIE Medical Imaging*, volume 6144, 2006. 3
- [33] W. M. Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, pages 846–850, 1971. 9
- [34] R. Reid and P. Scalzi. An improved colposcopic index for differentiating benign papillomavirus infections from high grade cervical intraepithelial neoplasia. *American Journal of Obstetrics and Gynecology*, 153(6):611–618, 1985. 2
- [35] M. Schiffman and P. Castle. Human papillomavirus. *Archives of Pathological Laboratory Medicine*, 127:930–934, 2003. 1
- [36] E. Seemann, B. Leibe, K. Mikolajczyk, and B. Schiele. An evaluation of local shape-based features for pedestrian detection. *Proc. BMVC*, 2005. 4
- [37] J. W. Sellors and R. Sankaranarayanan. *Colposcopy and Treatment of Cervical Intraepithelial Neoplasia: A Beginners Manual*. International Agency for Research on Cancer, 2003. 2
- [38] Y. Srinivasan, F. Gao, B. Tulpule, S. Yang, S. Mitra, and B. Nutter. Segmentation and classification of cervix lesions by texture and pattern analysis. *Intl. Journal of Intelligent Systems Technologies and Applications*, 1(3/4):234–246, 2006. 3
- [39] A. Staff. Cervicography: a new method for cervical cancer detection. *American Journal of Obstetrics and Gynecology*, 139(7):815–825, 1981. 1
- [40] L. Vincent and P. Soille. Watersheds in digital spaces: An efficient algorithm based on immersion simulations. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, pages 583–598, 1991. 4
- [41] Z. Xue, S. Antani, R. Long, and G. Thoma. Comparative performance analysis of cervix ROI extraction and specular reflection removal algorithms for uterine cervix image analysis. In *Proc. of SPIE Medical Imaging*, volume 6512, pages 1–9, 2007. 3
- [42] 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. 10
- [43] G. Zimmerman, S. Gordon, and H. Greenspan. Automatic landmark detection in uterine cervix images for indexing in a content-retrieval system. *Int. Symposium on Biomedical Imaging (ISBI)*, pages 1348–1351, 2006. 3