Unsupervised segmentation and classification of cervical cell images

Aslı Gençtav [a,1], Selim Aksoy [a,*], Sevgen Önder [b]

[a] Department of Computer Engineering, Bilkent University, Ankara 06800, Turkey
[b] Department of Pathology, Hacettepe University, Ankara 06100, Turkey

ARTICLE INFO

ABSTRACT

The Pap smear test is a manual screening procedure that is used to detect precancerous changes in cervical cells based on color and shape properties of their nuclei and cytoplasms. Automating this procedure is still an open problem due to the complexities of cell structures. In this paper, we propose an unsupervised approach for the segmentation and classification of cervical cells. The segmentation process involves automatic thresholding to separate the cell regions from the background, a multi-scale hierarchical segmentation algorithm to partition these regions based on homogeneity and circularity, and a binary classifier to finalize the separation of nuclei from cytoplasm within the cell regions. Classification is posed as a grouping problem by ranking the cells based on their feature characteristics modeling abnormality degrees. The proposed procedure constructs a tree using hierarchical clustering, and then arranges the cells in a linear order by using an optimal leaf ordering algorithm that maximizes the similarity of adjacent leaves without any requirement for training examples or parameter adjustment. Performance evaluation using two data sets show the effectiveness of the proposed approach in images having inconsistent staining, poor contrast, and overlapping cells.

1. Introduction

Cervical cancer is the second most common type of cancer among women with more than 250,000 deaths every year [1]. Fortunately, cervical cancer can be cured when early cancerous changes or precursor lesions caused by the Human Papilloma Virus (HPV) are detected. However, the cure rate is closely related to the stage of the disease at diagnosis time, with a very high probability of fatality if it is left untreated. Therefore, timely identification of the positive cases is crucial.

Since the discovery of a screening test, namely the Pap test, introduced by Dr. Georges Papanicolaou in the 1940s, a substantial decrease in the rate of cervical cancer and the related mortality was observed. The Pap test has been the most effective cancer screening test ever, and still remains the crucial modality in detecting the precursor lesions for cervical cancer. The test is based on obtaining cells from the uterine cervix, and smearing them onto glass slides for microscopic examination to detect HPV effects. The slides are stained using the Papanicolaou method where different components of the cells show different colors so that their examination becomes easier (see Figs. 1 and 2 for examples).

There are certain factors associated with the sensitivity of the Pap test, and thus, the reliability of the diagnosis. The sensitivity of the test is hampered mostly by the quality of sampling (e.g., number of cells) and smearing (e.g., presence of obscuring elements such as blood, mucus, and inflammatory cells, or poorly fixation of specimens). Both intra- and inter-observer variabilities during the interpretation of the abnormal smears also contribute to the wide variation in false-negative results [2]. The promise of early diagnosis as well as the associated difficulties in the manual screening process have made the development of automated or semi-automated systems that analyze images acquired by using a digital camera connected to the microscope an important research problem where more robust, consistent, and quantifiable examination of the smears is expected to increase the reliability of the diagnosis [3,4].

Both automated and semi-automated screening procedures involve two main tasks: segmentation and classification. Segmentation mainly focuses on separation of the cells from the background as well as separation of the nuclei from the cytoplasm within the cell regions. Automatic thresholding, morphological operations, and active contour models appear to be the most popular and common choices for the segmentation task in the literature. For example, Bamford and Lovell [5] segmented the nucleus in a Pap smear image using an active contour model that was estimated by using dynamic programming to find the boundary with the minimum cost within a bounded space around the darkest point in the image. Wu et al. [6] found the boundary

* Corresponding author. Tel.: +90 312 2903405; fax: +90 312 2664047.
E-mail addresses: asli@ceng.metu.edu.tr (A. Gençtav),
saksoy@cs.bilkent.edu.tr (S. Aksoy), sonder@hacettepe.edu.tr (S. Önder).
[1] Present address: Department of Computer Engineering, Middle East Technical University, Ankara 06531, Turkey.
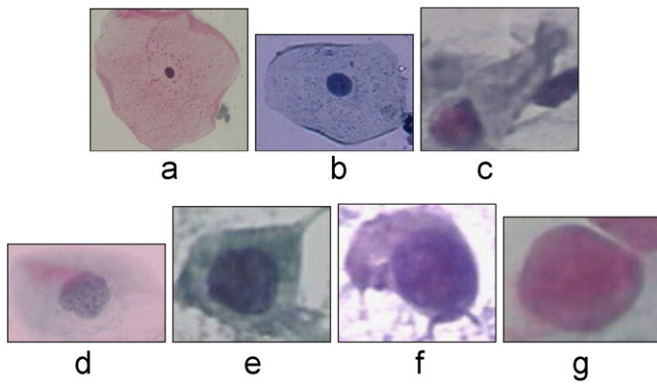
Fig. 1. Examples from the Herlev data set involving a single cell per image. The cells belong to (a) superficial squamous, (b) intermediate squamous, (c) columnar, (d) mild dysplasia, (e) moderate dysplasia, (f) severe dysplasia, and (g) carcinoma in situ classes. The classes in the first row are considered to be normal and the ones in the second row are considered to be abnormal. Average image size is 156 × 140 pixels. Details of this data set are given in Section 2.
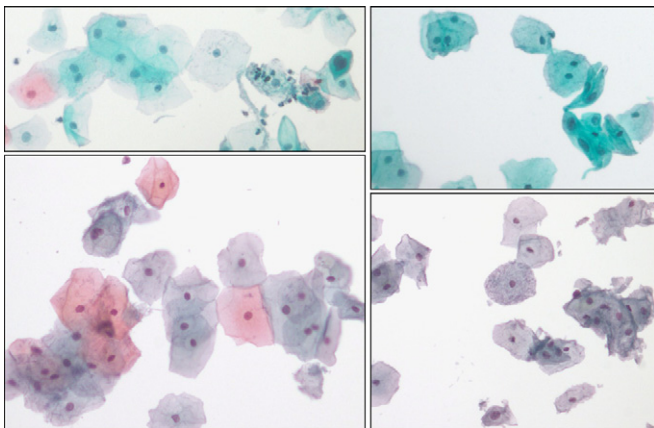

Fig. 2. Examples from the Hacettepe data set involving multiple overlapping cells with inconsistent staining and poor contrast that correspond to a more realistic and challenging setting. Details of this data set are given in Section 2.

of an isolated nucleus in a cervical cell image by using a parametric cost function with an elliptical shape assumption for the region of interest. Yang-Mao et al. [7] applied automatic thresholding to the image gradient to identify the edge pixels corresponding to nucleus and cytoplasm boundaries in cervical cell images. Tsai et al. [8] replaced the thresholding step with k-means clustering into two partitions. Dagher and Tom [9] combined the watershed segmentation algorithm with the active contour model by using the watershed segmentation result of a down-sampled image as the initial contour of the snake for the segmentation of blood cells and corneal cells. Huang and Lai [10] also used the marker-based watershed segmentation algorithm to find an approximate segmentation, applied heuristic rules to eliminate non-nuclei regions, and used active contours to improve the nuclei boundaries in biopsy images of liver cells. Harandi et al. [11] used the active contour algorithm to identify the cervical cell boundaries, applied thresholding to identify the nucleus within each cell, and then used a separate active contour for each nucleus to identify the corresponding cytoplasm within connected cell groups. Plissiti et al. [12] detected the locations of nuclei centroids in Pap smear images by using the local minima of image gradient, eliminated the candidate centroids that were too close to each other, and used a support vector machine (SVM) classifier for the final selection of points using color values in

square neighborhoods. Then, they used the detected centroids as markers in marker-based watershed segmentation to find the nuclei boundaries, and eliminated the false-positive regions by using a binary SVM classifier with shape, texture, and intensity features [13]. Li et al. [14] used k-means clustering for a rough partitioning of an image into nucleus, cytoplasm, and background areas, and then performed snake-based refinement of nucleus and cytoplasm boundaries.

Most of these methods focus on the segmentation of only the nuclei [5,6,9,10,12,13] for which there is relatively higher contrast around the boundaries. However, detection of the cytoplasm regions is also crucial because cytoplasm features have been shown to be very useful for the identification of abnormal cells [15]. Even so, these nuclei-specific methods do not necessarily generalize well for the detection of cytoplasms that create an increased difficulty due to additional gradient content and local variations. Furthermore, many of the methods [5–8,14] assume a single cell in the input image where there is only one boundary (nucleus) or at most two boundaries (nucleus and cytoplasm) to detect as in the examples in Fig. 1. However, this is usually not a realistic setting as can be seen in the images in Fig. 2 where one cannot make any assumption about the number of cells or expect that these cells appear isolated from each other so that they can be analyzed independently. Among the proposed methods, automatic thresholding for nuclei detection [7,8] assumes a bimodal distribution but can only be used for the segmentation of isolated cells. Watershed-based methods [10,12,13] can identify more details and can handle multiple cells but have the potential of over-segmentation, so they require carefully adjusted preprocessing steps or carefully selected markers. Active contour-based methods [5,9–11,14] can provide a better localization of boundaries when there is sufficient contrast but are often very difficult to initialize with a very sensitive process for parameter and capture range selections. Our earlier work showed that it can be very difficult to find reliable and robust markers for marker-based watershed segmentation and very difficult to find a consistent set of parameters for active contour models when there are multiple cells in the image [16]. A recent development of interest has been the work on the incorporation of shape priors into active contour models to resolve overlapping and occluded objects [17]. However, these priors have been mainly applied to the segmentation of objects with a well-defined and consistent appearance, whereas it is not straightforward to define a shape prior for the overlapping cells with highly varying cytoplasm areas as shown in Fig. 2. Moreover, their initialization is still an existing problem when the number of cells and their approximate locations are unknown.

In this paper, our first major contribution is a generic and parameter-free segmentation algorithm that can delineate cells and their nuclei in images having inconsistent staining, poor contrast, and overlapping cells. The first step in the segmentation process separates the cell regions from the background using morphological operations and automatic thresholding that can handle varying staining and illumination levels. Then, the second step builds a hierarchical segmentation tree by using a multi-scale watershed segmentation procedure, and automatically selects the regions that maximize a joint measure of homogeneity and circularity with the goal of identifying the nuclei at different scales. The third step finalizes the separation of nuclei from cytoplasm within the segmented cell regions by using a binary classifier. The proposed algorithms are different from related work in that (1) the automatic thresholding step can handle multiple cell groups in images because the gray scale bimodality assumption holds when the goal is to extract only the background, and (2) no initialization, parameter adjustment, or marker detection are required. Unlike some other approaches that tried to select a single scale from watershed hierarchies [18] or use thresholds on region features to select a subset of regions
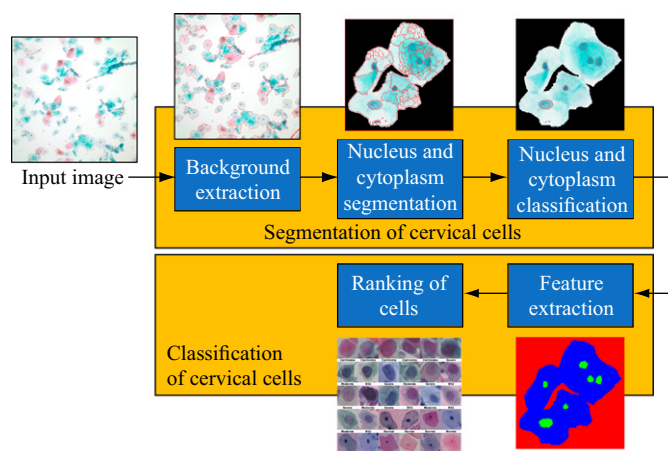
Fig. 3. Overview of the approach. The segmentation step aims to separate the cell regions from the background and identify the individual nuclei within these regions. The classification step provides an unsupervised ranking of the cells according to their degree of abnormality.

from watershed segmentations performed at multiple scales [19,20], the proposed algorithm automatically selects the meaningful regions from the hierarchical segmentation tree built by using local characteristics in multiple scales. Furthermore, the whole procedure is unsupervised except the final nucleus versus cytoplasm classification step. Our main focus in this paper is to correctly identify the individual nuclei under the presence of overlapping cells while assuming that the overlapping cytoplasm areas are shared by different cells in the rest of the analysis. Segmentation of individual cytoplasm areas needs future research because we have observed that an accurate delineation of the cytoplasm region for each cell in our image spatial resolution may not be realistic in the presence of heavily overlapping cells with inconsistent staining and poor contrast. Fig. 3 provides an overview of the proposed approach.

After segmentation, classification mainly focuses on automatic labeling of the cells into two classes: normal versus abnormal. For example, Walker et al. [21] used a quadratic Gaussian classifier with co-occurrence texture features extracted from the nucleus pixels. Chou and Shapiro [22] classified cells using more than 300 features with a hierarchical multiple classifier algorithm. Zhang and Liu [23] performed pixel-based classification using 4000 multispectral features with SVM-based feature selection. Marinakis et al. [15] used 20 features computed from both nucleus and cytoplasm regions using feature selection with a genetic algorithm and a nearest neighbor classifier. They also considered a more detailed seven-class problem but observed a decrease in accuracy. These studies showed that high success rates can be obtained with various classification methods particularly for the normal versus abnormal cell identification in controlled data sets. However, all of these methods require a large number of example patterns for each class, but sampling a sufficient number of training data from each of the classes is not always possible. Furthermore, since these classifiers require complete descriptions of all classes, they may not generalize well with a sufficiently high accuracy especially for diverse data sets having large variations in appearance within both normal and abnormal categories. An alternative mode of operation that is also of interest in this paper is to quantify the abnormality of the cells and present a restricted set of views to the human expert for semi-automatic analysis [24]. This approach has the potential to reduce screening errors and increase the throughput, since the manual screening can allocate more time on cells that are more likely to be abnormal.

As our second major contribution, we propose an unsupervised approach for the grouping of the cells into multiple classes.

Even though supervised classification has been the main focus in the literature, we avoid using any labels in this paper due to the potential practical difficulties of collecting sufficient number of representative samples in a multi-class problem with unbalanced categories having significantly different observation frequencies in a realistic data set. However, the number and distribution of these groupings are also highly data dependent because instances of some classes may not be found in a particular image, with the extreme being all normal cells. In this paper, we pose the grouping problem as the ranking of cells according to their degree of abnormality. The proposed ranking procedure, first, constructs a binary tree using hierarchical clustering according to the features extracted from the nucleus and cytoplasm regions. Then, an optimal leaf ordering algorithm arranges the cells in a linear order by maximizing the similarity of adjacent leaves in the tree. The algorithm provides an automatic way of organizing the cells without any requirement for training examples or parameter adjustment. The algorithm also enables an expert to examine the ranked list of cells and evaluate the extreme cases in detail while skipping the cells that are ranked as more normal than a selected cell that is manually confirmed to be normal.

The rest of the paper is organized as follows. The data sets used for illustrating and evaluating the proposed algorithms are described in Section 2. The algorithm for the segmentation of cells from the background is presented in Section 3. The segmentation procedure for the separation of nuclei from cytoplasm is discussed in Section 4. The procedure for unsupervised classification of cells via ranking is described in Section 5. Quantitative and qualitative performance evaluation are presented in Section 6. Conclusions are given in Section 7.

2. Data sets

The methodologies presented in this paper are illustrated using two different data sets. The first one, the Herlev data set, consists of 917 images of single Pap cells, and was collected by the Department of Pathology at Herlev University Hospital and the Department of Automation at Technical University of
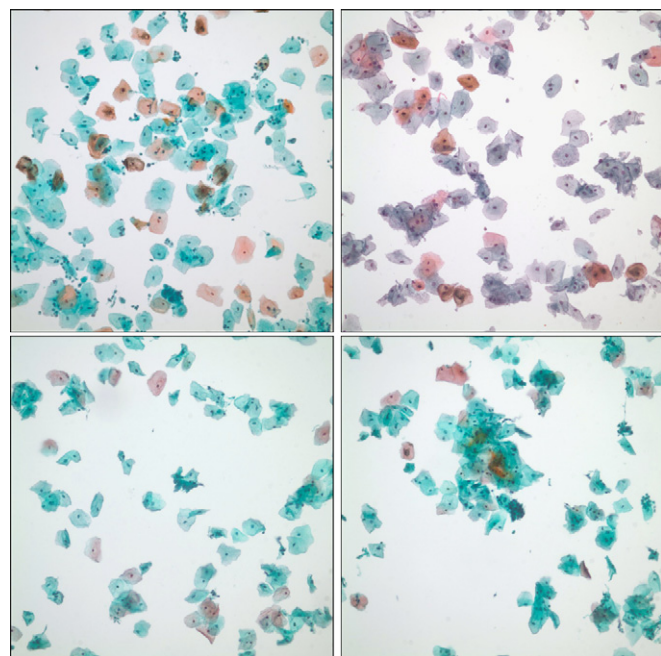

Fig. 4. Examples of full images from the Hacettepe data set. Each image has 2048×2048 pixels.

Denmark [25]. The images were acquired at a magnification of 0.201 μm/pixel. Average image size is 156×140 pixels. Cyto-technicians and doctors manually classified each cell into one of the seven classes, namely superficial squamous, intermediate squamous, columnar, mild dysplasia, moderate dysplasia, severe dysplasia, and carcinoma in situ. The first three classes correspond to normal cells and the last four classes correspond to abnormal cells with examples in Fig. 1. Each cell also has an associated ground truth of nucleus and cytoplasm regions.

The second data set, referred to as the Hacettepe data set, was collected by the Department of Pathology at Hacettepe University Hospital using the ThinPrep liquid-based cytology preparation technique. It consists of 82 Pap test images belonging to 18 different patients. Each image has 2048×2048 pixels, and was acquired at 200X magnification. These images are more realistic with the challenges of overlapping cells, poor contrast, and inconsistent staining with examples shown in Fig. 4. We manually delineated the nuclei in a subset of this data set for performance evaluation. Both data sets are used for quantitative and qualitative evaluation in this paper.

3. Background extraction

The background extraction step aims at dividing a Pap smear test image into cell and background regions where cell regions correspond to the regions containing cervical cells and background regions correspond to the remaining empty area. As a result of the staining process, cell regions are colored with tones of blue and red whereas background regions remain colorless and produce white pixels. We observed that cell and background regions have distinctive colors in terms of brightness, and consequently, they can be differentiated according to their lightness. To that end, we first separate color and illumination information by converting Pap smear test images from the original RGB color space to the CIE Lab color space, and then analyze the L channel representing the brightness level. Fig. 5(a)–(c) illustrate an image and its L channel with the corresponding histogram. The particular choice of the L channel is further explained at the end of this section.

An important factor is that Pap smear test images usually have the problem of inhomogeneous illumination due to uneven lightening of the slides during image acquisition. We correct the inhomogeneous illumination in the L channel by using the black top-hat transform. The black top-hat, also known as top-hat by closing [26], of an image I is the difference between the closing of the image and the image itself, and is computed as

$$\text{BTH}(I,SE) = (I \bullet SE) - I \tag{1}$$

where SE is the structuring element that is selected as a disk larger than the largest connected component of cells. A radius of 210 pixels is empirically selected in this paper. Since the cells are darker than the background, the black top-hat produces an evenly illuminated image, called the illumination-corrected L channel, with cells that are brighter than the background after subtraction. Fig. 5(d)–(f) illustrates the correction process.

After illumination correction, the separation of the cells from the background can be formulated as a binary detection problem using a threshold. Even though one cannot set a threshold a priori for all images due to variations in staining, it is possible to assume a bimodal distribution where one mode corresponds to the background and the other mode corresponds to the cell regions so that automatic thresholding can be performed. We use minimum-error thresholding [27] to separate these modes automatically. We assume that the respective populations of background and cell regions have Gaussian distributions. Given the histogram of the L

channel as $H(x), x \in [0, L_{\max}]$, that estimates the probability density function of the mixture population and a particular threshold value $T \in (0, L_{\max})$, the two Gaussians can be estimated as

$$P(w_i|T) = \sum_{x=a}^{b} H(x) \tag{2}$$

$$p(x|w_i,T) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(x-\mu_i)^2}{2\sigma_i^2}\right) \tag{3}$$

where w_1 and w_2 correspond to the background and cell regions, respectively, and

$$\mu_i = \frac{1}{P(w_i|T)} \sum_{x=a}^{b} x H(x) \tag{4}$$

$$\sigma_i^2 = \frac{1}{P(w_i|T)} \sum_{x=a}^{b} (x-\mu_i)^2 H(x) \tag{5}$$

with $\{a,b\} = \{0,T\}$ for w_1 and $\{a,b\} = \{T+1, L_{\max}\}$ for w_2. Then, the optimal threshold can be found as the one that minimizes the criterion function

$$J(T) = -\sum_{x=0}^{L_{\max}} H(x) P(w_i|x,T), \quad i = \begin{cases} 1, & x \leq T \\ 2, & x > T \end{cases} \tag{6}$$

where $P(w_i|x,T)$ for $i = 1, 2$ are computed from (2) and (3) using the Bayes rule. Fig. 5(g) shows the histogram of the illumination-corrected L channel in 5(f), and the criterion function $J(T)$ for different values of T is shown in Fig. 5(h). Fig. 5(i) shows the result for the optimal threshold found as 0.03. We also analyzed the histograms and thresholding results obtained by using other color spaces. Results are not given here due to space constraints but we observed that the L channel gave the most consistent bimodal distribution in our empirical evaluation.

4. Segmentation of cervical cells

Segmentation of cell regions into nucleus and cytoplasm is a challenging task because Pap smear test images usually have the problems of inconsistent staining, poor contrast, and overlapping cells. We propose a two-phase approach for segmentation of cell regions that can contain single cells or many overlapping cells. The first phase further partitions the cell regions by using a non-parametric hierarchical segmentation algorithm that uses spectral and shape information as well as gradient information. The second phase identifies nucleus and cytoplasm regions by classifying the segments resulting from the first phase by using multiple spectral and shape features. Parts of this section were presented in [28].

4.1. Nucleus and cytoplasm segmentation

The first phase aims at partitioning each connected component of cell regions into a set of sub-regions where each nucleus is accurately represented by a segment while the rest of the segments correspond to parts of cytoplasms. We observed that it may not be realistic to expect an accurate segmentation of individual cytoplasm regions for each cell in this resolution in the presence of overlapping cells with inconsistent staining and poor contrast. Therefore, the proposed segmentation algorithm focuses on obtaining each nucleus accurately while the remaining segments are classified as cytoplasm in the second phase.

4.1.1. Hierarchical region extraction

The main source of information that we use to delineate nuclei regions is the relative contrast between these nuclei and cytoplasm regions. The watershed segmentation algorithm is an
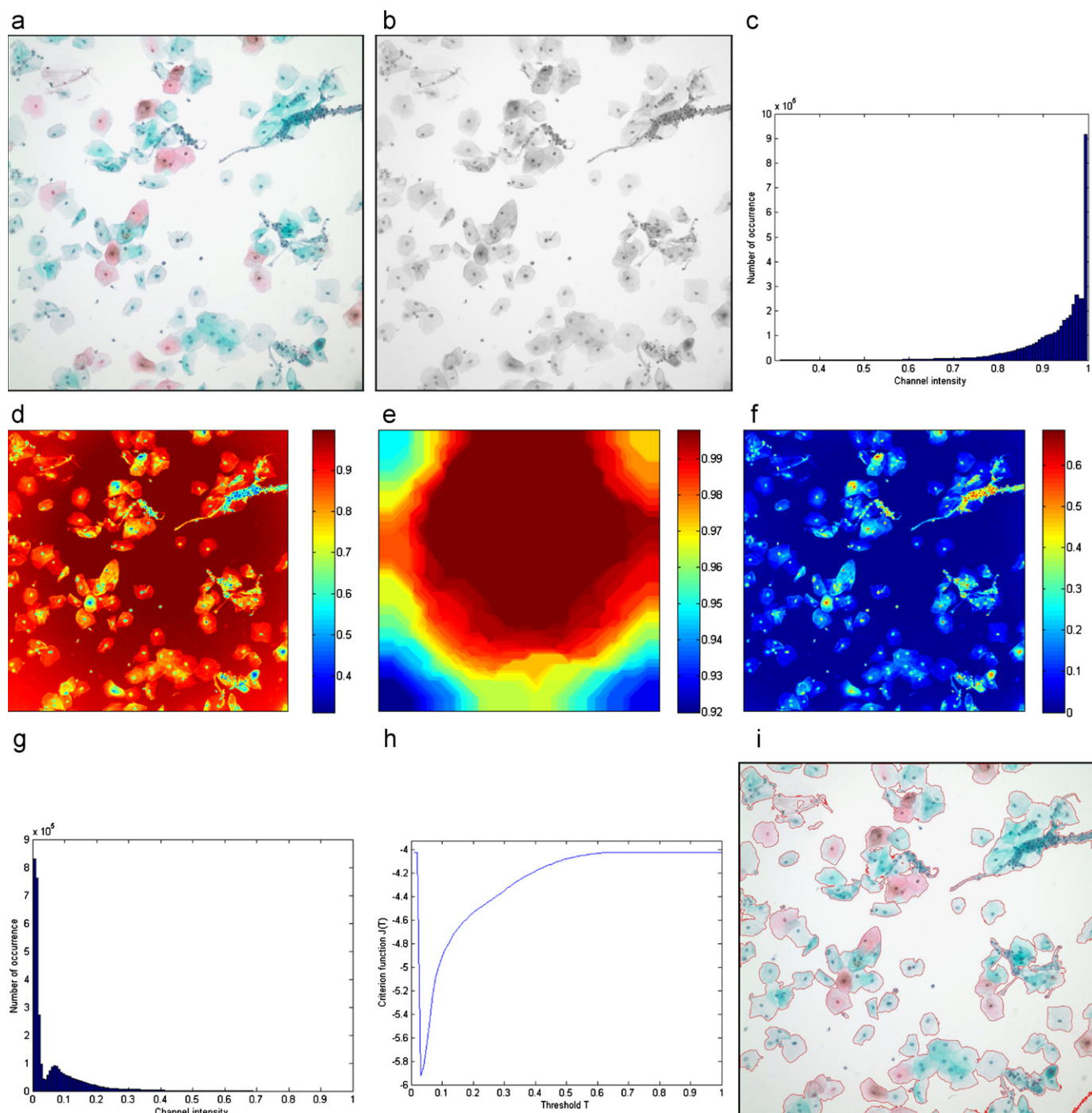
Fig. 5. Background extraction example. (a) Pap smear image in RGB color space. (b) L channel of the image in CIE Lab color space. (c) Histogram of the L channel. (d) L channel shown in pseudo color to emphasize the contrast. (e) Closing with a large structuring element. (f) Illumination-corrected L channel in pseudo color. (g) Histogram of the illumination-corrected L channel. (h) Criterion for automatic thresholding. (i) Results of thresholding at 0.03 with cell region boundaries marked in red. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

effective method that models the local contrast differences using the magnitude of image gradient with the additional advantage of not requiring any prior information about the number of segments in the image. However, watersheds computed from raw image gradient often suffer from over-segmentation. Furthermore, it is also very difficult to select a single set of parameters for pre- or post-processing methods for simplifying the gradient so that the segmentation is equally effective for multiple structures of interest. Hence, we use a multi-scale approach to allow

accurate segmentation under inconsistent staining and poor contrast conditions.

We use multi-scale watershed segmentation that employs the concept of dynamics that are related to regional minima of image gradient [29] to generate a hierarchical partitioning of cell regions. A regional minimum is composed of a set of neighboring pixels with the same value x where the pixels on its external boundary have a value greater than x. When we consider the image gradient as a topographic surface, the dynamic of a regional

minimum can be defined as the minimum height that a point in the minimum has to climb to reach a lower regional minimum.

The multi-scale watershed segmentation generates a set of nested partitions of a cell region. The partition at scale s is obtained as the watershed segmentation of the image gradient whose regional minima with dynamics less than or equal to s are
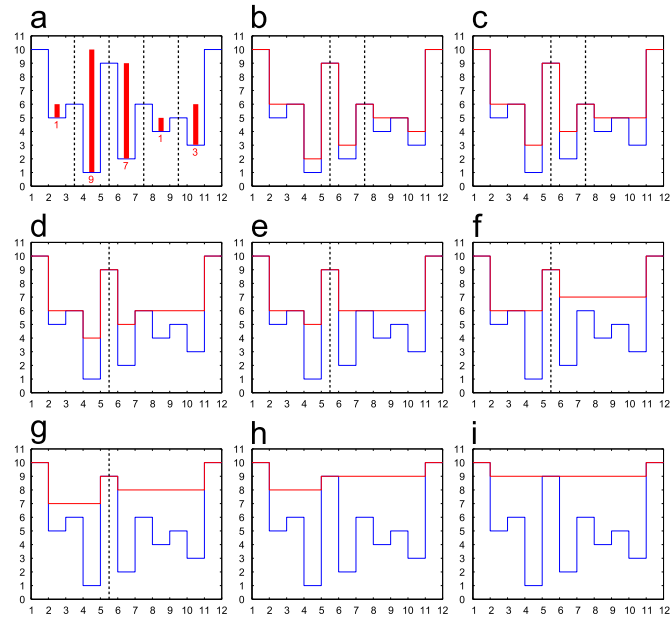

Fig. 6. One-dimensional synthetic signal (blue) and watersheds (black) of h-minima transforms (red) at different scales. The dynamic of each initial regional minimum is also shown as red bars in (a). In all figures, the y-axis simulates the signal values (e.g., image gradient) and the x-axis simulates the domain of the signal (e.g., pixel locations). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

eliminated using the h-minima transform. The h-minima transform suppresses all minima having dynamics less than or equal to h by performing geodesic reconstruction by erosion of the input image f from $f+h$ [26]. Fig. 6 illustrates the multi-scale watershed segmentation on a one-dimensional synthetic signal. The first partition is calculated as the classical watershed with five catchment basins corresponding to five local minima. The two catchment basins having a dynamic of 1 are merged with their neighbor catchment basins at scale 1. At each scale s, the minima with dynamics less than or equal to s are filtered whereas the minima with dynamics greater than s remain the same or are extended. This continues until the last scale corresponding to the largest dynamic in the gradient image. Thus, the range of scales starts at scale 0 that corresponds to the raw image gradient, and ends at the largest dynamic with a maximum value of 255, enabling automatic construction of the scales for each image based on its gradient content. Fig. 7 illustrates the segmentation process on a cell region at six different scales. The minima of the raw image gradient mainly mark the texture occurring in nuclei and cytoplasm. More regional minima are filtered as the scale increases, and a correct segment for each nucleus is obtained at some scale because the nucleus segments are associated with higher dynamic values.

A hierarchical tree can be constructed from the multi-scale partitions of a cell region if we ensure that the partitions are nested, i.e., a segment in a partition of a certain scale either remains the same or is contained in a larger segment at the next scale. An important point is that the watershed segmentation that uses image gradients smoothed using the h-minima transform does not always satisfy this nesting property. For example, the nested structure of the partitions is disturbed when the gradient image has a minimum similar to the one in the middle in Fig. 8(a). The middle segment at scale 0 is split into two and merged with different segments at the next scale in Fig. 8(b), because the watershed line between the two regional minima at scale 1 is found at its center. The watershed lines are adjusted so that a segment splitting at the next scale is merged
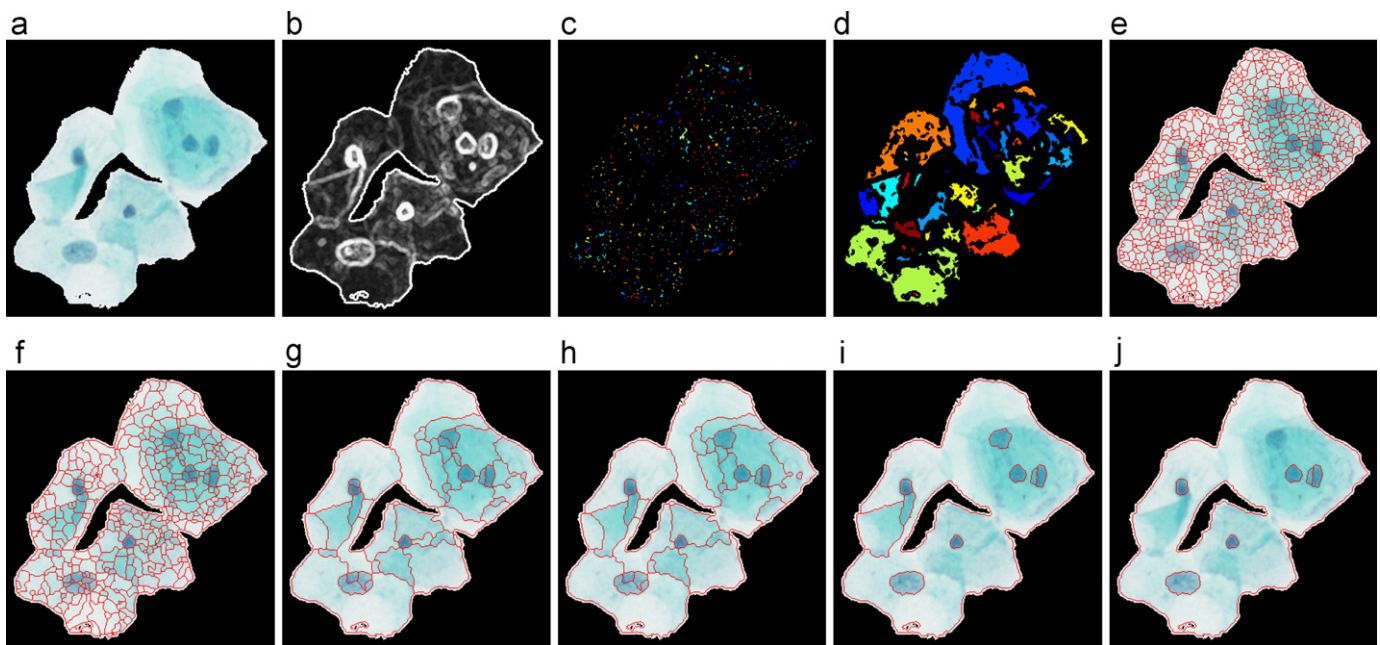

Fig. 7. Hierarchical region extraction example. (a) An example cell region (the removed background is shown as black). (b) Gradient of the L channel that is used for computing the dynamics. (c) Regional minima of the raw gradient shown in pseudo color. (d) Minima obtained after the h-minima transform for $h=6$. Candidate segments obtained by multi-scale watershed segmentation at (e) scale 0, (f) scale 1, (g) scale 6, (h) scale 7, (i) scale 13, and (j) scale 14 with boundaries marked in red. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)
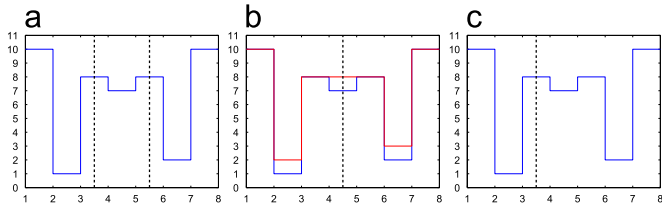
Fig. 8. One-dimensional synthetic signal (blue) and watersheds (black) of h-minima transforms (red) at (a) scale 0 and (b) scale 1. (c) The partition at scale 1 after the proposed adjustment. In all figures, the y-axis simulates the signal values (e.g., image gradient) and the x-axis simulates the domain of the signal (e.g., pixel locations). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

with its neighbor segment having the most similar mean intensity value. We illustrate this solution in Fig. 8(c) as the middle segment at scale 0 is merged with its right neighbor at scale 1 assuming that the mean intensity of the middle segment and its right neighbor are more similar.

After ensuring the nested structure of the partitions, we construct a hierarchical tree from all segments of each scale where each segment is a node and there is an edge between two nodes of consecutive scales if one node is contained within the other. Thus, the leaf nodes are the segments obtained from the watershed segmentation of the raw gradient image, and the root becomes the whole cell region. Fig. 9 demonstrates an example tree for several scales for a real image.

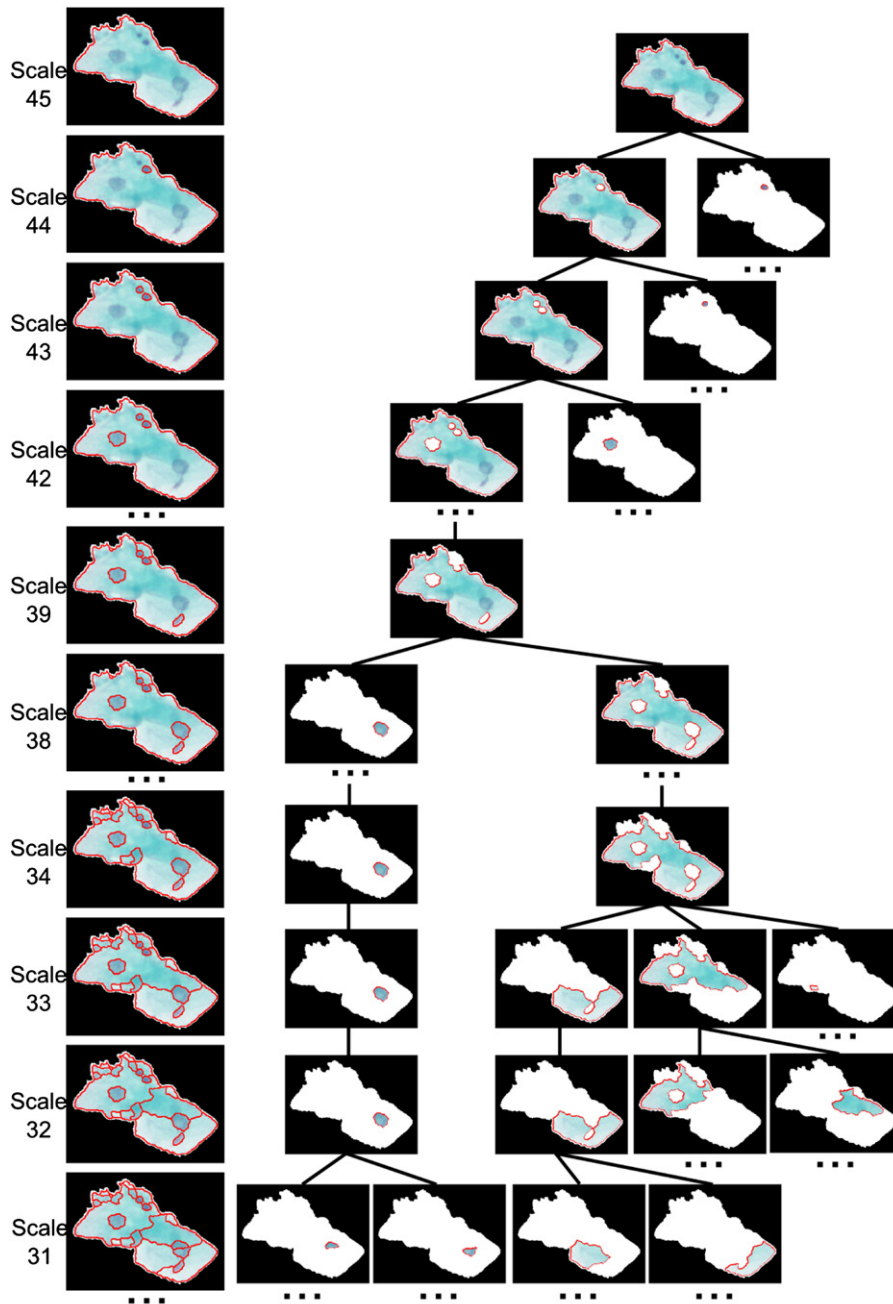**Fig. 9.** Example tree for segmentation hierarchy. Segmentations at example scales are shown on the left column. Parts of the corresponding tree are illustrated on the right. Nodes correspond to the segments and the edges represent the containment relation between the segments in two consecutive scales. A particular segment in a certain scale either remains the same or merges with other segments to form a larger segment.
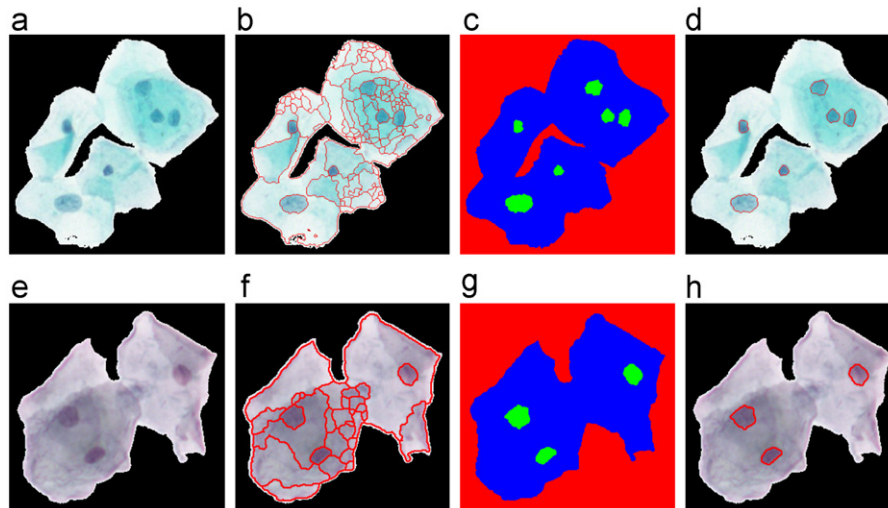
Fig. 10. Region selection example. (a), (e) Example cell regions. (b), (f) Region selection results. (c), (g) Classification results shown in pseudo color: background (red), nucleus (green), and cytoplasm (blue). (d), (h) Resulting nucleus boundaries marked in red. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

4.1.2. Region selection

Each node of the hierarchical tree is regarded as a candidate segment for the final segmentation. Our aim is to select the most meaningful segments among those appearing at different levels of the tree. Nucleus regions are considered the most meaningful segments, because their appearances are expected to be more consistent compared to the appearance of the cytoplasm regions, and they can be differentiated according to their spectral homogeneity and shape features in the resolution at which we operate.

In general, small segments in lower levels of the hierarchy merge to form nucleus or cytoplasm regions in higher levels of the hierarchy where homogeneous and circular nucleus regions are obtained at some level. These nucleus regions may stay the same for some number of levels, and then, face a large change at a particular scale because they merge with their surrounding segments of cytoplasm. The segments that we are interested in are the homogeneous and circular regions right before this change. Thus, the goodness measure of a node is calculated in terms of two factors: homogeneity and circularity.

The homogeneity measure of a node R_1 is determined based on its spectral similarity to its parent node R_2, and is quantified using the F-statistic

$$F(R_1, R_2) = \frac{(n_1 + n_2 - 2)n_1 n_2}{n_1 + n_2} \frac{(m_1 - m_2)^2}{s_1^2 + s_2^2}, \tag{7}$$

where n_i is the number of pixels, m_i is the mean of the pixels, and s_i^2 is the scatter of the pixels belonging to R_i for $i = 1, 2$. The F-statistic measures the significance of the difference of the means of two distributions relative to their pooled variance [30,31]. In this context, a small F-value is observed when a node remains the same or merges with similar regions in the next level, indicating an insignificant difference between their means. On the other hand, a large F-value implies that the node merges with regions with different spectral features; thus, disturbing the homogeneity of the node in the next level, and resulting in a large difference of the means. The F-statistic in (7) requires each pixel of R_1 and R_2 to be associated with a single value. Hence, the spectral values of each pixel in the original three-dimensional RGB space are projected onto a line using Fisher's linear discriminant analysis [32], and the projected values are used to compute (7).

The circularity measure C of a node R is defined as the multiplicative inverse of eccentricity e [30] of the ellipse that has the same second moments as the corresponding segment of that node, i.e.,

$$C(R) = 1/e. \tag{8}$$

The circularity measure is maximized for a segment that has a circular shape because eccentricity is minimum for a circle and is maximum for a line segment.

Finally, the goodness measure Γ of a node R is defined as

$$\Gamma(R) = F(R, parent(R)) \times C(R). \tag{9}$$

Given the goodness measure of each node in the hierarchical tree, the segments that optimize this measure are selected by using a two-pass algorithm that we previously proposed for the segmentation of remotely sensed images [33]. Given $\mathcal{N}$ as the set of all nodes and $\mathcal{P}$ as the set of all paths in the tree, the algorithm selects $\mathcal{N}^* \subseteq \mathcal{N}$ as the final segmentation such that any node in $\mathcal{N}^*$ must have a measure greater than all of its descendants, any two nodes in $\mathcal{N}^*$ cannot be on the same path (i.e., the corresponding segments cannot overlap in the hierarchical segmentation), and every path must include a node that is in $\mathcal{N}^*$ (i.e., the segmentation must cover the whole image). The first pass finds the nodes having a measure greater than all of their descendants in a bottom-up traversal. The second pass selects the most meaningful nodes having the largest measure on their corresponding paths of the tree in a top-down traversal. The details of the algorithm can be found in [33]. As shown in Fig. 10, the final segmentation contains all of the true nucleus regions and several sub-regions belonging to the cytoplasm as the most meaningful segments.

4.2. Nucleus and cytoplasm classification

The second phase aims to produce a final partitioning of the cell regions into nucleus and cytoplasm regions by classifying the resulting segments of the previous phase. The classification of each segment as nucleus or cytoplasm is based on multiple spectral and shape features, namely, size, mean intensity, circularity, and homogeneity.

The data set used for training and evaluating the classifiers consists of 1452 nucleus regions and 7726 cytoplasm regions

Table 1
Classification of segments as nucleus or cytoplasm. The number of misclassified nuclei (N) out of 726, the number of misclassified cytoplasms (C) out of 3863, and the total number of misclassified segments (T) out of 4589 are used as evaluation criteria.

	Classifier	N	C	T
1	Bayesian	38	216	254
2	Decision tree	96	86	182
3	Support vector machine	99	50	149
4	Combination using sum	71	80	151
5	Combination using product	65	96	161

manually labeled from different cell regions in the Hacettepe data set. While collecting data from a cell region, all of the nucleus and cytoplasm regions resulting from the segmentation of that cell region were gathered in order to preserve the class frequencies. After partitioning the data set into equally sized training and validation sets, the performances of different classifiers were evaluated using the four features that were normalized to the [0, 1] range by linear scaling.

The classification performances of different classifiers are given in Table 1. The first classifier is a Bayesian classifier that uses multivariate Gaussians for class-conditional densities and class frequencies for prior probabilities. The second classifier is a decision tree classifier built by using information gain as the binary splitting criterion and a pessimistic error estimate for pruning. The third one is an SVM classifier using the radial basis function kernel. We also combined these three classifiers using sum and product of individual posterior probabilities.

Even though the SVM classifier had the best performance in terms of the overall accuracy, we chose the combined classifier based on the sum of posterior probabilities, because it had a higher accuracy for the classification of nucleus regions. Fig. 10 shows classification results for example cell images. The combination of segmentation and classification results show that the four features and the trained classifiers accurately identify the nucleus regions with an overall correct classification rate of 96.71%. The final cytoplasm area for each cell region is obtained by taking the union of all segments classified as cytoplasm within that particular region.

5. Classification of cervical cells

As discussed in Section 1, the cell classification problem is defined here as an unsupervised grouping problem. Different from many other unsupervised clustering approaches, the proposed procedure does not make any assumption about the distribution of the groups. It also does not require any information regarding the number of groups in the data. Given the motivation of identifying problematic cells as regions of interest for expert assistance, we pose the grouping problem as the ranking of cells according to their abnormality degrees. Ranking has been an important problem in pattern recognition and information retrieval where the patterns are ordered based on their similarity to a reference pattern called the query. However, the ranking methods in the literature are not directly applicable to our problem because it does not involve any query cell. In this section, we propose an unsupervised non-parametric ordering procedure that uses a tree structure formed by hierarchical clustering. First, we present the features that are used for describing the segmented cells. Then, we describe the details of the ordering algorithm.

5.1. Feature extraction

Dysplastic changes of cervical cells can be associated with cell characteristics like size, color, shape, and texture of nucleus and cytoplasm. We describe each cell by using 14 different features related to these characteristics. The extracted features are a subset of the features used in [25] for cervical cells.

A cell region may contain a single cell or several overlapping cells. In the latter case, the segmentation result consists of an individual nucleus for each cell and a single cytoplasm region that is the union of overlapping cytoplasms of all cells. Since each cell has a single nucleus, the number of cells in an overlapping cell region is equal to the number of nuclei segments found in that cell region. Consequently, we approximate the association of the cytoplasm to each individual cell within a group of overlapping cells by distributing an equal share to each cell based on the number of nuclei. Then, a set of features is extracted for each nucleus and the shared cytoplasm as:

- *Nucleus area*: The number of pixels in the nucleus region.
- *Nucleus brightness*: The average intensity of the pixels belonging to the nucleus region.
- *Nucleus longest diameter*: The diameter of the smallest circle circumscribing the nucleus region. We calculate it as the largest distance between the boundary pixels that form the maximum chord of the nucleus region.
- *Nucleus shortest diameter*: The diameter of the largest circle that is totally encircled by the nucleus region. It is approximated by the length of the maximum chord that is perpendicular to the maximum chord computed above.
- *Nucleus elongation*: The ratio of the shortest diameter to the longest diameter of the nucleus region.
- *Nucleus roundness*: The ratio of the nucleus area to the area of the circle corresponding to the nucleus longest diameter.
- *Nucleus perimeter*: The length of the perimeter of the nucleus region.
- *Nucleus maxima*: The number of pixels that are local maxima inside a 3×3 window.
- *Nucleus minima*: The number of pixels that are local minima inside a 3×3 window.
- *Cytoplasm area*: The number of pixels in the cytoplasm part of a cell region divided by the number of cells in that cell region. We assume that the total cytoplasm is shared equally by the cells in a cell region.
- *Cytoplasm brightness*: Calculated similar to the nucleus brightness. However, overlapping cells are associated with the same value of the cytoplasm brightness.
- *Cytoplasm maxima*: Calculated similar to the nucleus maxima feature. Overlapping cells are associated with the same value.
- *Cytoplasm minima*: Calculated similar to the nucleus minima feature. Overlapping cells are associated with the same value.
- *Nucleus/cytoplasm ratio*: This feature measures how small the nucleus of a cell is compared to its cytoplasm. It is given by the ratio of the nucleus area to the cell area which is calculated as the sum of the nucleus and cytoplasm area.

5.2. Ranking of cervical cells

We use hierarchical clustering to produce a grouping of cells according to the features described above. Hierarchical clustering constructs a binary tree in which each cell corresponds to an individual cluster in the leaf level, and the two most similar clusters merge to form a new cluster in the subsequent levels. The clusters that are merged are selected based on pairwise distances in the form of a distance matrix. We use the Euclidean distance for

computing the pairwise feature distances and the average linkage criterion to compute the distance between two clusters [32].

Hierarchical clustering and the resulting tree structure are intuitive ways of organizing the cells because the cells that are adjacent in the tree are assumed to be related with respect to their feature characteristics. These relations can be converted to a linear ordering of the cells corresponding to the ordering of the leaf nodes. Let T be a binary tree with n leaf nodes denoted as $z_1,\ldots,z_n$, and $n-1$ non-leaf nodes denoted as $y_1,\ldots,y_{n-1}$. A linear ordering that is consistent with T is defined to be an ordering of the leaves of T that is generated by flipping the non-leaf nodes of T, i.e., swapping the left and right subtrees rooted at y_i for any $y_i \in T$ [34]. A flipping operation at a particular node changes the order of the subtrees of that node, and produces a different ordering of the leaves. Fig. 11 illustrates the flipping of subtrees at a node where the ordering of the leaves is changed while the same tree structure is preserved. The possibility of applying a flipping operation at each of the $n-1$ non-leaf nodes of T results in a total of 2^{n-1} possible linear orderings of the leaves of T.

Fig. 12 shows an example binary tree (dendrogram) generated as a result of hierarchical clustering of 30 cells consisting of randomly selected five cells each from six classes in the Herlev data set. As can be seen from this tree, the dysplastic cells are first organized into nested clusters, and then the clusters of normal cells are formed. The clusters of dysplastic and normal cells are later merged into a single cluster. The leaf ordering for this particular visualization of the generated hierarchical grouping uses a combination of pairwise distance values and indices of the cells in the data in ascending order. Fig. 13(a) shows the cells and

their class names corresponding to this ordering. The dysplastic cells are found at the beginning of the ordering and the normal cells are grouped at the end of the list. However, the sub-classes of dysplastic cells are not accurately ordered according to their dysplasia degree, and the group of normal cells at the end of the list contains some dysplastic cells as well. The main reason is that many heuristic orderings only consider local associations and do not consider any global consistency.

It is possible to compute an optimal leaf ordering for a given tree where optimality is defined as the maximum sum of similarities of adjacent leaves in the ordering. Given the space Φ of all 2^{n-1} possible orderings of the leaves of T, the goodness $D^{\phi}(T)$ of a particular ordering $\phi \in \Phi$ can be defined as

$$D^{\phi}(T) = \sum_{i=1}^{n-1} S(z_i^{\phi}, z_{i+1}^{\phi}) \tag{10}$$

where z_i^{ϕ} is the i'th leaf when T is ordered according to ϕ, and S is the pairwise similarity matrix. Bar-Joseph et al. [34] described an algorithm for finding the ordering that maximizes (10). The algorithm runs in $O(n^4)$ time, and uses dynamic programming by recursively computing the goodness of the optimal ordering of the subtree rooted at each non-leaf node y in a bottom up way. The worst case running time of this algorithm remains at $O(n^4)$ for balanced binary trees, but the computation time dramatically decreases on average for less balanced trees generated using hierarchical clustering of cell features.

The measure in (10) for the goodness of an ordering in terms of the sum of similarities between adjacent leaves can be modified as the sum of similarities between every leaf and all other leaves in the adjacent clusters for a more global agreement. The adjacent clusters of a particular leaf node z are found as follows. If z is on the left (right) branch of its parent, then all leaf nodes that belong to the right (left) subtree of its parent are considered as the right (left) adjacent cluster of z. To find the left (right) adjacent cluster of z, we go up to the ancestors of z until we reach an ancestor that has a left (right) subtree that does not contain z, and all leaf nodes that belong to that subtree are considered as the left (right) adjacent cluster of z. For example, in Fig. 11(a), the left adjacent cluster of the leaf 3 contains the leaves 1 and 2, and its right adjacent cluster consists of the leaves 4 and 5. Hence, the set of the similarities between the leaf 3 and its adjacent clusters becomes $\{S(3,1), S(3,2), S(3,4), S(3,5)\}$. As a result, the goodness measure for this particular ordering of the tree can be calculated as the sum of the pairwise similarities in the union of the sets
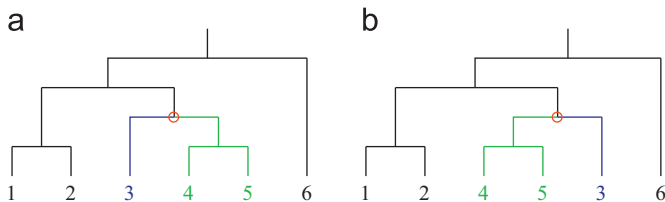
a b

Fig. 11. Leaf ordering example. (a) An example binary tree T with the leaf nodes labeled with integers from 1 to 6. (b) A leaf ordering consistent with T obtained by flipping the node surrounded by the red circle. The flipping operation at a node corresponds to swapping the left subtree and the right subtree of that node. The left and right subtrees of the flipped node in (a) are shown in blue and green, respectively. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)
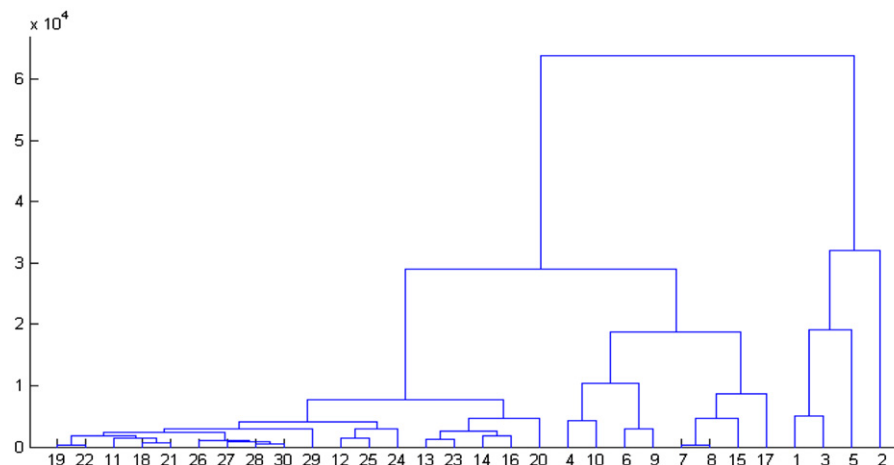

Fig. 12. The binary tree resulting from hierarchical clustering of 30 cells randomly selected from the Herlev data (normal superficial (1–5), normal intermediate (6–10), mild dysplasia (11–15), moderate dysplasia (16–20), severe dysplasia (21–25), carcinoma in situ (26–30)).
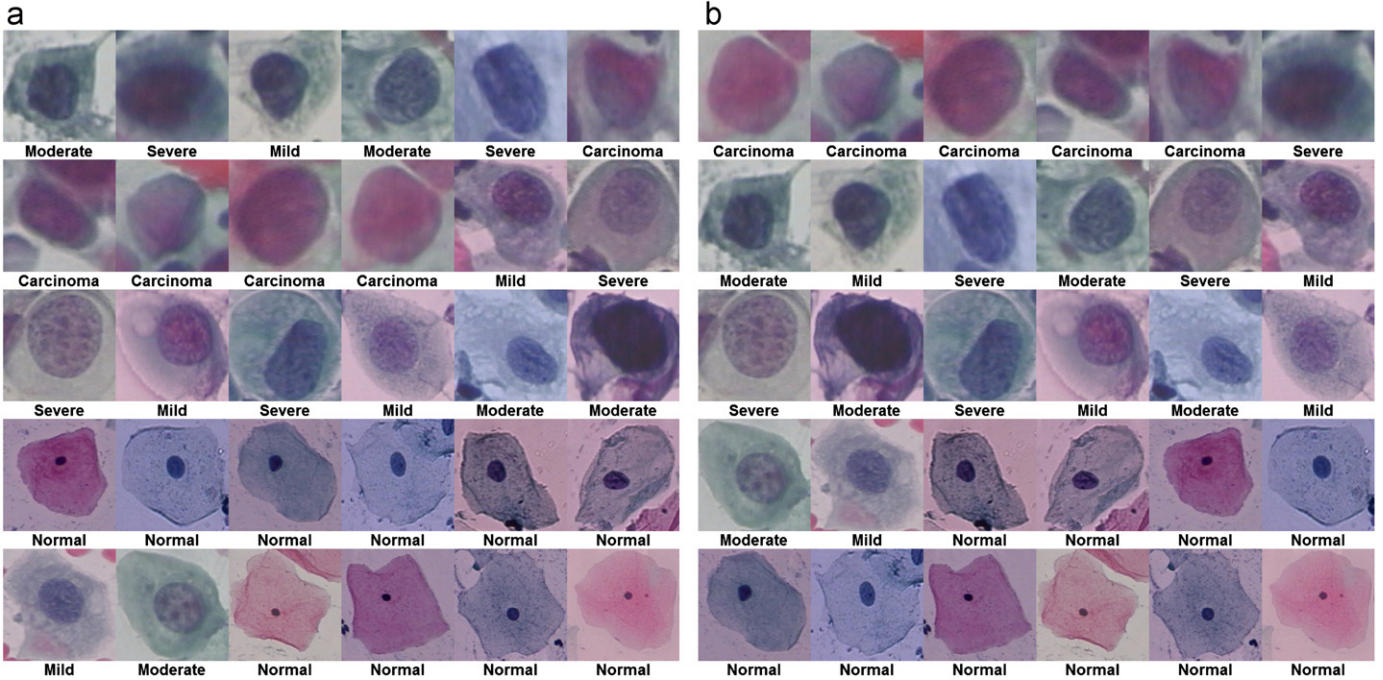
a

b

Fig. 13. Example orderings of cells. (a) Initial ordering of the cells determined by the linear ordering of the leaves of the original tree in Fig. 12. (b) Final ordering obtained by applying the optimal leaf ordering algorithm.

$\{S(1,2)\}$, $\{S(2,1)$, $S(2,3)$, $S(2,4)$, $S(2,5)\}$, $\{S(3,1)$, $S(3,2)$, $S(3,4)$, $S(3,5)\}$, $\{S(4,3)$, $S(4,5)\}$, $\{S(5,4)$, $S(5,6)\}$, and $\{S(6,1)$, $S(6,2)$, $S(6,3)$, $S(6,4)$, $S(6,5)\}$. This measure can be formally defined as

$$D^{\phi}(\mathcal{T}) = \sum_{i=1}^{n} \sum_{j \in A_i^{\phi}} S(z_i^{\phi}, z_j^{\phi}) \qquad (11)$$

where A_i^{ϕ} is the set of nodes in the adjacent clusters of z_i^{ϕ}. We use both of the measures (10) and (11) in the ranking of the cells by using the optimal leaf ordering algorithm on the binary tree obtained by hierarchical clustering. Similarity S is computed as the additive inverse (negative) of the distance between cell pairs. Fig. 13(b) shows the cells and their class names corresponding to the optimal ordering of the leaves in the tree shown in Fig. 12. The result indicates that the optimal ordering can improve the ordering in Fig. 13(a).

6. Experimental results

The performance of our proposed computer-assisted screening methodology was evaluated using the Herlev and Hacettepe data sets described in Section 2. Except the nucleus and cytoplasm classification step that is described in Section 4.2 (which is supervised), our methods are fully automated and unsupervised so that there is no need to set a parameter or to collect training data. The experiments for the evaluation of background extraction, segmentation, and classification are presented below.

6.1. Evaluation of background extraction

The Herlev data set consists of single cell images many of which do not have any background area so it is not informative to evaluate background extraction on this data set. However, there is no ground truth data involving the boundaries between the cell regions and the background area in the images in the Hacettepe data. Therefore, in Fig. 14, we give illustrative examples covering a wide range of the images existing in the Hacettepe data to evaluate the performance of this phase qualitatively. The histogram of the illumination-corrected L channel is also given for each example.

The first example image in Fig. 14 comprises a large number of cell regions with many overlapping cells. After filtering the non-homogeneous illumination, we obtained a bimodal histogram. We can observe that the background was smoothly extracted for this Pap test image consisting of many cells. The second example illustrates an image with a smaller number of cells. The cell regions were also extracted accurately, although there were also two false detections. These false cell regions were due to the intensive non-homogeneous illumination of this Pap smear image. The false cell region with an oval shape in the center of the image was also affected from the pollution in that part of the slide. Compared to the other two examples, the cell density in the last image is in between the corresponding cell densities of the first and second images. The cell regions in this image were colored with different color tones compared to the previous images due to inconsistent staining. The background extraction performed well on this image except the false cell region detected at the bottom-right corner. The performance for most of the images in the Hacettepe data set resembles the first example in Fig. 14, with possible error cases shown in the second and third examples in Fig. 14. Overall, the background extraction phase performs well under varying conditions, and is a very practical method based on automatic thresholding. It may only suffer from the intensive non-homogeneous illumination especially at the image corners. The uneven illumination of the images arises from the image acquisition stage which can be improved using a better controlled setup to overcome this problem.

6.2. Evaluation of segmentation

The performance of our segmentation procedure for locating nucleus regions with a correct delineation of their boundaries was compared against the manually constructed ground truth. Since
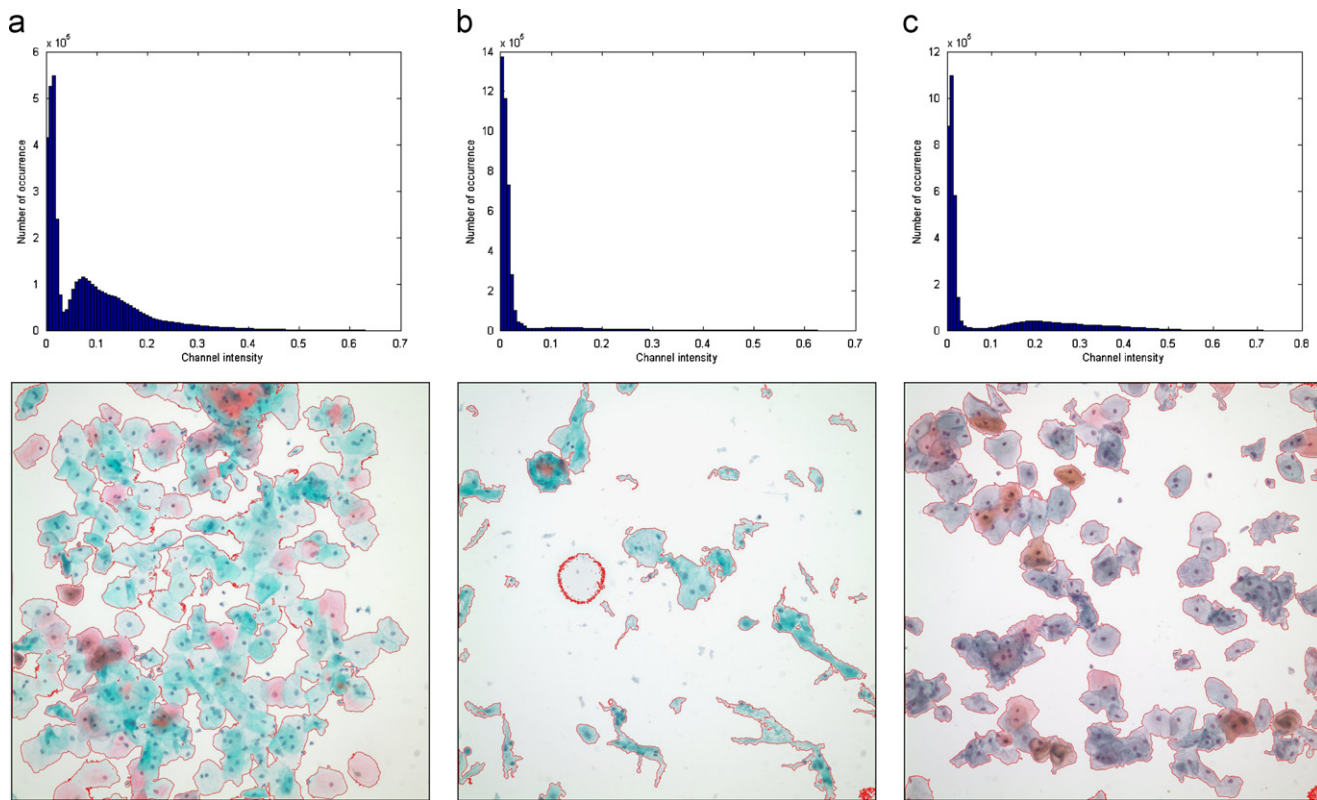
Fig. 14. Background extraction examples. The histogram of the illumination-corrected L channel and the corresponding result for thresholding are given with cell boundaries marked in red. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

the Herlev data set consisted of single cell images, a one-to-one correspondence could be found between the segmentation results and the ground truth nuclei by choosing the segment that had the highest overlap with the ground truth nucleus region for comparison. However, since the Hacettepe images consisted of multiple cells, a matching step was required to find correspondences between the resulting segments and the ground truth nuclei. We used an object-based evaluation procedure similar to [35] that was adapted from the work of [36] on range image segmentation evaluation. This procedure used the individual reference objects in the ground truth and the output objects in the produced segmentation, and classified every pair of reference and output objects as correct detections, over-detections, under-detections, missed detections, or false alarms with respect to a threshold τ on the amount of overlap between these objects. The overlap was computed in terms of number of pixels. A pair of reference and output objects was classified as an instance of correct detection if at least τ percent of each object overlapped with the other. A reference object and a set of output objects were classified as an instance of over-detection if at least τ percent of each output object overlapped with the reference object and at least τ percent of the reference object overlapped with the union of the output objects. An output object and a set of reference objects were classified as an instance of under-detection if at least τ percent of each reference object overlapped with the output object and at least τ percent of the output object overlapped with the union of the reference objects. A reference object that was not in any instance of correct detection, over-detection, and under-detection was classified as missed detection. An output object that was not in any instance of correct detection, over-detection, and under-detection was classified as false alarm. An overlap threshold of $\tau = 0.6$ was used in the experiments in this paper. Once all reference and output nuclei were classified into instances of correct detections, over-detections, under-detections, missed

detections, or false alarms, we computed object-based precision and recall as the quantitative performance criteria as

$$\text{precision} = \frac{\text{\# of correctly detected objects}}{\text{\# of all detected objects}} = \frac{N - FA}{N} \qquad (12)$$

$$\text{recall} = \frac{\text{\# of correctly detected objects}}{\text{\# of all objects in the ground truth}} = \frac{M - MD}{M} \qquad (13)$$

where FA and MD were the number of false alarms and missed detections, respectively, and N and M were the number of nuclei in the segmentation output and in the ground truth, respectively.

The accuracies of the detected segmentation boundaries were also quantified using pixel-based precision and recall. Evaluation was performed using nuclei pairs, one from the segmentation output and the other one from the ground truth, that were identified as correct detection instances as described above. For each pair, the number of true-positive (TP), false-positive (FP), and false-negative (FN) pixels were counted to compute precision and recall as

$$\text{precision} = \frac{\text{\# of correctly detected pixels}}{\text{\# of all detected pixels}} = \frac{TP}{TP + FP} \qquad (14)$$

$$\text{recall} = \frac{\text{\# of correctly detected pixels}}{\text{\# of all pixels in the ground truth}} = \frac{TP}{TP + FN} \qquad (15)$$

We also computed the Zijdenbos similarity index (ZSI) [37] that is defined as the ratio of twice the common area between two regions A_1 and A_2 to the sum of individual areas as

$$\text{ZSI} = 2\frac{|A_1 \cap A_2|}{|A_1| + |A_2|} = \frac{2TP}{2TP + FP + FN}, \qquad (16)$$

resulting in ZSI $\in [0, 1]$. This similarity index, also known as the Dice similarity coefficient in the literature, considers differences in both size and location where an ZSI greater than 0.7 indicates an excellent agreement between the regions [37].

We compared our results to those of a state-of-the-art active contour-based segmentation algorithm designed for cervical cell images [14]. The algorithm started with a rough partitioning of an image into nucleus, cytoplasm, and background regions using k-means clustering, and obtained the final set of boundaries using radiating gradient vector flow (RGVF) snakes. Although the original algorithm was designed for single cell images, we applied it to the Hacettepe data set by initializing a separate contour for each connected component of the nucleus cluster of the k-means result. The publicly available code cited in [14] was used in the experiments. For the Herlev data set, the segment having the highest overlap with the single-cell ground truth was used as the initial contour for the RGVF method. For the Hacettepe data set, the default parameters in [14] were employed, except the area threshold used in initial contour extraction. Instead of eliminating the regions whose areas were smaller than a fixed percentage of the whole image, we eliminated the candidate nucleus regions whose areas were smaller than 100 pixels. It might be possible to obtain better results for some cases by tuning the parameters but we kept the default values as it was not possible to find a single set of parameter values that worked consistently well for all images.

Table 2 presents the results of pixel-based evaluation for the Herlev data set. The average precision, recall, and ZSI measures were all close to 0.9 for the proposed algorithm. We believe that this is a very satisfactory result as our algorithm did not use any assumption about the size, location, or the number of cells in an image even though the Herlev data set consists of single cell images. The RGVF algorithm also achieved similar overall results, but it took advantage of the knowledge of the existence of a single cell in each image through accordingly designed preprocessing and initialization procedures.

The superiority of the proposed algorithm became apparent in the experiments that used the Hacettepe data set. Table 3 presents the results of object-based evaluation. It can be seen that the proposed algorithm obtained significantly higher number of correct detections along with significantly lower missed detections and lower false alarms. These statistics led to significantly higher precision and recall rates compared to the RGVF algorithm. The results showed the power of the generic nature of the proposed algorithm that could handle images containing overlapping cells without any requirement for initialization or parameter adjustment. On the other hand, we observed that the RGVF algorithm had difficulties in obtaining satisfactory results for a consistent set of parameters for different images in this challenging data set. The algorithm designed for single-cell images was not easily generalizable to these realistic images in which the additional local extrema caused by the overlapping cytoplasms of multiple cells and the folding cytoplasm of individual cells caused severe problems.

Table 4 presents the results of pixel-based evaluation for the Hacettepe data set. Both the proposed algorithm and the RGVF algorithm achieved similar results in terms of average precision, recall, and ZSI measures for the nuclei that were identified as correctly detected during the object-based evaluation. The results for the RGVF algorithm were actually relatively higher than those for the proposed algorithm but it is important to note that these averages were computed using only a small number of relatively easy cells that could be detected by the RGVF algorithm (using only 51 and 47 detected nuclei for RGVF settings 1 and 2, respectively, as shown in Table 3), whereas the results for the proposed algorithm were obtained from much higher number of cells (using 130 detected nuclei).

Figs. 15 and 16 show example results from the Hacettepe data set. For each example cell region in Fig. 16, a group of three images were given to show the ground truth nuclei, the results of the proposed algorithm, and the results of the RGVF algorithm with using area-based elimination. It could be observed that the RGVF algorithm easily got stuck on local gradient variations or missed the nuclei altogether due to poor initializations. There were also

Table 3
Object-based evaluation of segmentation results for the Hacettepe data set. The number of ground truth nuclei (M), output nuclei (N), correct detections (CD), over-detections (OD), under-detections (UD), missed detections (MD), and false alarms (FA), as well as precision (prec) and recall (rec) are given for both the proposed algorithm and the RGVF algorithm. RGVF setting 1 corresponds to elimination of regions smaller than 100 pixels during initialization, and setting 2 corresponds to an additional elimination of regions that are not round enough as in [14].

Algorithm	M	N	CD	OD	UD	MD	FA	prec	rec
Proposed	139	174	130	0	0	9	44	0.7471	0.9353
RGVF setting 1	139	127	51	0	1	86	75	0.4095	0.3813
RGVF setting 2	139	63	47	0	0	92	16	0.7460	0.3381

Table 4
Pixel-based evaluation of segmentation results for the Hacettepe data set. The mean μ and standard deviation σ of precision, recall, and the Zijdenbos similarity index (ZSI) computed only for the nuclei identified as correctly detected in the object-based evaluation (using 51 and 47 nuclei for RGVF settings 1 and 2, respectively, and 130 nuclei for the proposed algorithm as shown in Table 3) are given for both the proposed algorithm and the RGVF algorithm.

Algorithm	$\mu_{prec} \pm \sigma_{prec}$	$\mu_{rec} \pm \sigma_{rec}$	$\mu_{ZSI} \pm \sigma_{ZSI}$
Proposed	0.91 + 0.08	0.88 + 0.07	0.89 + 0.04
RGVF setting 1	0.95 + 0.06	0.89 + 0.08	0.91 + 0.04
RGVF setting 2	0.95 + 0.06	0.89 + 0.08	0.91 + 0.04

Table 2
Pixel-based evaluation of segmentation results for the Herlev data set. The mean μ and standard deviation σ of precision, recall, and the Zijdenbos similarity index (ZSI) for each class of the Herlev data are given for both the proposed algorithm and the RGVF algorithm.

|	Class name	Class size	Proposed algorithm			RGVF algorithm		
			$\mu_{prec} \pm \sigma_{prec}$	$\mu_{rec} \pm \sigma_{rec}$	$\mu_{ZSI} \pm \sigma_{ZSI}$	$\mu_{prec} \pm \sigma_{prec}$	$\mu_{rec} \pm \sigma_{rec}$	$\mu_{ZSI} \pm \sigma_{ZSI}$
Normal	Superficial squamous	74 cells	0.69 + 0.37	0.63 + 0.37	0.98 + 0.12	0.92 + 0.12	0.88 + 0.14	0.98 + 0.02
	Intermediate squamous	70 cells	0.79 + 0.29	0.73 + 0.31	0.98 + 0.12	0.95 + 0.03	0.92 + 0.06	0.98 + 0.02
	Columnar	98 cells	0.85 + 0.15	0.77 + 0.18	0.98 + 0.05	0.83 + 0.16	0.76 + 0.20	0.97 + 0.08
Abnormal	Mild dysplasia	182 cells	0.88 + 0.17	0.86 + 0.16	0.96 + 0.16	0.92 + 0.13	0.90 + 0.16	0.96 + 0.08
	Moderate dysplasia	146 cells	0.91 + 0.10	0.86 + 0.14	0.97 + 0.07	0.89 + 0.15	0.87 + 0.17	0.94 + 0.13
	Severe dysplasia	197 cells	0.90 + 0.12	0.89 + 0.11	0.95 + 0.13	0.88 + 0.15	0.90 + 0.13	0.90 + 0.19
	Carcinoma in situ	150 cells	0.89 + 0.15	0.90 + 0.08	0.92 + 0.17	0.84 + 0.18	0.88 + 0.11	0.86 + 0.24
	Average	917 cells	0.88 + 0.15	0.93 + 0.15	0.89 + 0.15	0.83 + 0.20	0.96 + 0.13	0.87 + 0.19
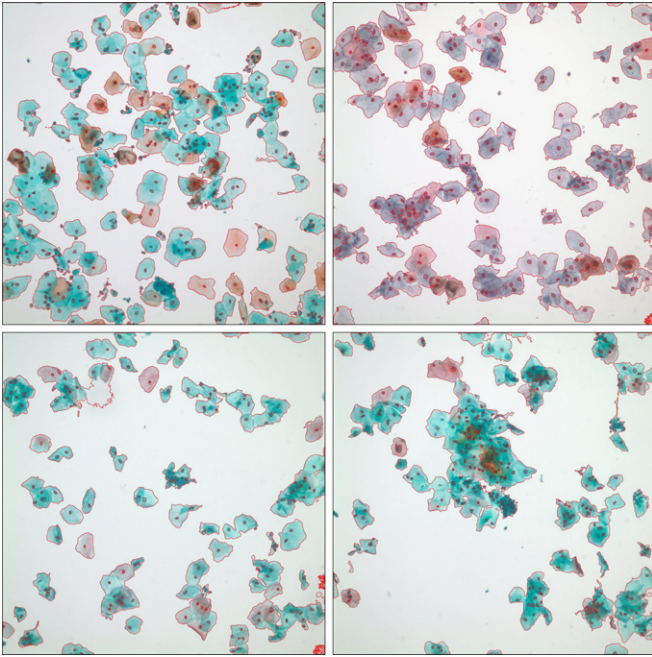
Fig. 15. Segmentation results for full images from the Hacettepe data set. The boundaries of the cell regions and the nuclei found within these regions using the proposed algorithm are marked in red. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

some occasions in which segments of nucleus regions could not be obtained using our method. In some cases, a nucleus region never appeared in the hierarchy due to its noisy texture or insufficient contrast with the surrounding cytoplasm. These cases may occur when the camera is out of focus for these particular cells, or when the nuclei overlap with the cytoplasms of other cells. Capturing images at multiple focus settings may allow the detection of additional nuclei [38]. In some other cases, even though a nucleus appeared in the hierarchical tree, its ancestor at a higher level was found to be more meaningful because of the selection heuristics, or the selected nucleus was later misclassified as cytoplasm. However, our method is generic in the sense that it allows additional measures for defining the meaningfulness of a segment to be employed (through additional terms in (9) in Section 9) and alternative methods for the classification of segments to be used (through additional classifiers in Section 4.2), so that the results can be further improved. Post-processing steps can also be designed to eliminate false alarms resulting from inflammatory cells that may appear in the algorithm output because they also have dark round shapes but are not in the ground truth because they are not among the cervical cells of interest.

6.3. Evaluation of classification

We use the following experimental protocol to evaluate the performance of unsupervised classification using ranking. The Herlev data set is used because the cells have ground truth class labels. First, a set of I cells belonging to the Herlev data are ranked according to their class labels. This ranking corresponds to the ideal one that we would like to achieve because our goal is to order the cells according to their abnormality degrees. In this way, we obtain the ground truth ranking U where we know the rank U_i of each cell $q_i, i = 1, \ldots, I$. An example set of cells $q_1, q_2, q_3, q_4, q_5, q_6, q_7, q_8$ belonging to three different classes are given in Table 5. The ranks of the cells with the same class label should be the same so we assign all of these cells to the mean of

their initial ranks and obtain the ground truth ranking U. Then, suppose that our algorithm ranks these cells in the order of $q_1, q_5, q_2, q_4, q_3, q_7, q_6, q_8$. Since we aim to order the cells according to their abnormality degrees, we can hypothesize that our method labels the first three cells, namely q_1, q_5, q_2, as class 1, the next three cells, namely q_4, q_3, q_7, as class 2, and the last two cells, namely q_6, q_8, as class 3 because the classes 1, 2, and 3 are known to have three, three, and two images, respectively, in the ground truth. When we calculate the cell rankings based on these class associations, we obtain the ranking result V shown in Table 5 where each cell q_i has a corresponding rank $V_i, i = 1, \ldots, I$. Finally, we measure the agreement between the ground truth ranking U and our ranking result V statistically using the Spearman rank-order correlation coefficient and the kappa coefficient. These statistics and the corresponding results are given below.

The Spearman rank-order correlation coefficient R_s is defined as

$$R_s = \frac{\sum_{i=1}^{I}(U_i - \overline{U})(V_i - \overline{V})}{\sqrt{\sum_{i=1}^{I}(U_i - \overline{U})^2}\sqrt{\sum_{i=1}^{I}(V_i - \overline{V})^2}} \tag{17}$$

where $\overline{U}$ and $\overline{V}$ are the means of U_i's and V_i's, respectively. The sign of R_s denotes the direction of the correlation between U and V. If R_s is zero, then V does not increase or decrease while U increases. When U and V are highly correlated, the magnitude of R_s increases.

Unlike simple percent agreement, the kappa coefficient also considers the agreement occurring by chance. Suppose that two raters label each of the I observations into one of K categories. We obtain a confusion matrix N where N_{ij} represents the number of observations that are labeled as category i by the first rater and category j by the second rater. We also define a weight matrix W where a weight $W_{ij} \in [0, 1]$ denotes the degree of similarity between two categories i and j. The weights on the diagonal of W are selected as 1, whereas the weights W_{ij} with highly different categories i and j are determined to be close or equal to 0. The weighted relative observed agreement among raters is obtained as

$$P_o = \frac{1}{I} \sum_{i=1}^{K} \sum_{j=1}^{K} W_{ij} N_{ij}. \tag{18}$$

The weighted relative agreement expected just by chance is estimated by

$$P_e = \frac{1}{I^2} \sum_{i=1}^{K} \sum_{j=1}^{K} W_{ij} r_i c_j \tag{19}$$

where $r_i = \sum_{j=1}^{K} N_{ij}$ and $c_j = \sum_{i=1}^{K} N_{ij}$. Then, the weighted kappa coefficient κ_w which may be interpreted as the chance-corrected weighted relative agreement is given by

$$\kappa_w = \frac{P_o - P_e}{1 - P_e} \tag{20}$$

When all categories are equally different from each other, we obtain Cohen's kappa coefficient kappa by setting the weights W_{ij} in (18) and (19) to 0 for $i \neq j$. Both kappa coefficients have the maximum value of 1 when the agreement between the raters is perfect whereas the result is 0 in the case of no agreement.

We use the weight matrix

$$W = \begin{pmatrix} 1 & 0.5 & 0 & 0.25 & 0.25 & 0 & 0 \\ 0.5 & 1 & 0 & 0.25 & 0.25 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0.25 & 0.25 & 0 & 1 & 0.5 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0 & 0.5 & 1 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0.25 & 0.5 & 1 & 0.5 \\ 0 & 0 & 0 & 0.25 & 0.5 & 0.5 & 1 \end{pmatrix} \tag{21}$$

to compute the weighted kappa coefficient κ_w. The rows and columns correspond to the classes in Fig. 1 and Table 2. The non-zero
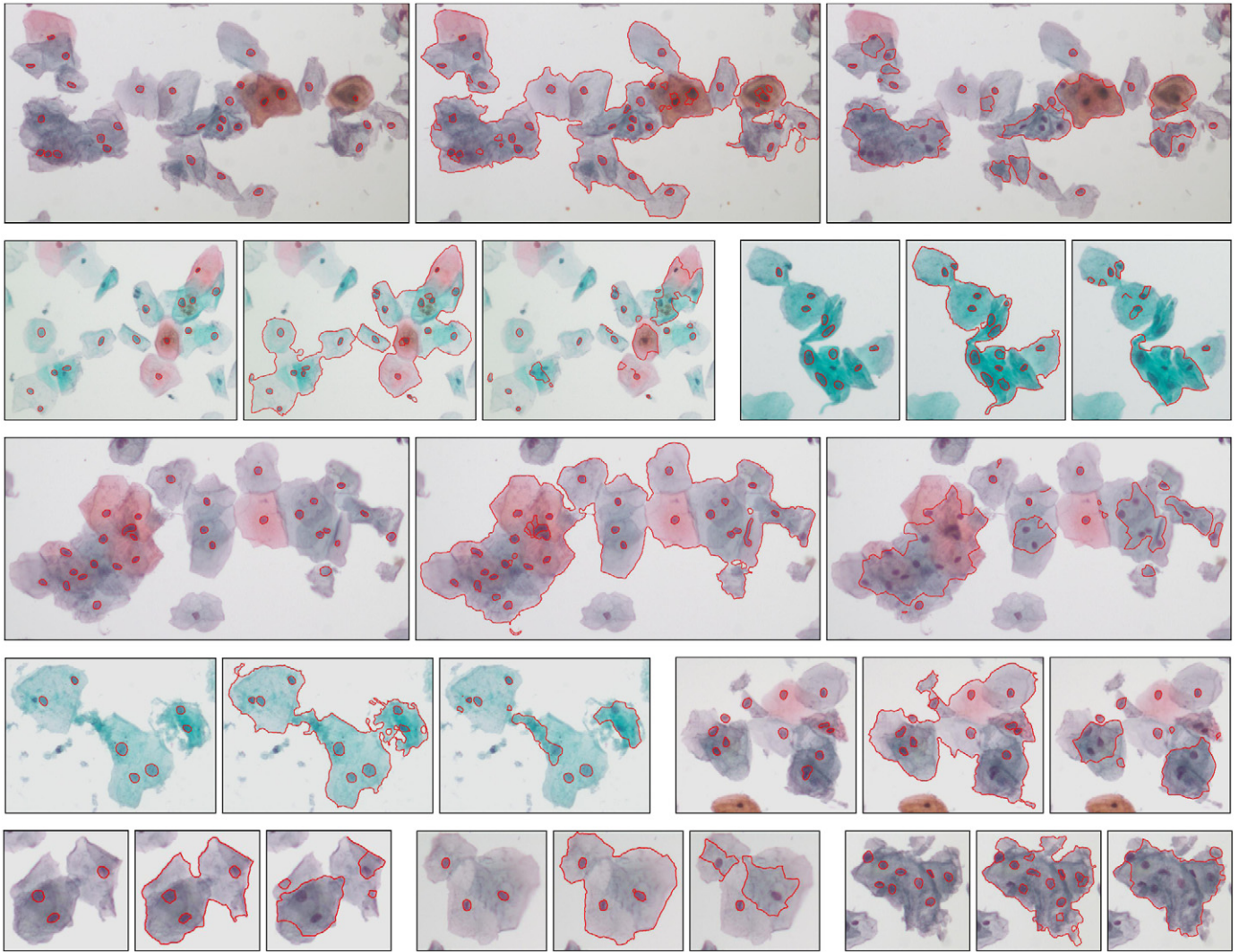
Fig. 16. Segmentation results for example images from the Hacettepe data set. For each group of three images, the ground truth nuclei, the result of the proposed segmentation algorithm, and the result of the RGVF algorithm with setting 1 are given. The resulting region boundaries are marked in red. The results for the proposed algorithm also include the boundaries of the detected cell regions in addition to the segmented nuclei. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

Table 5
An example ranking scenario. Eight cells are assumed to belong to three classes. The ground truth ranking U and the algorithm's ranking V are calculated according to the scenario described in the text.

Cells	q_1	q_2	q_3	q_4	q_5	q_6	q_7	q_8
Class labels	1	1	1	2	2	2	3	3
Initial ranking	1	2	3	4	5	6	7	8
Ground truth ranking U	2	2	2	5	5	5	7.5	7.5
Algorithm ranking V	2	2	5	5	2	7.5	5	7.5

Table 6
Ranking results for different settings of the cervical cell classes. Higher values of R_s, κ, and κ_w indicate better performance. The classes used for each case are marked using shaded rectangles.

	Classes used							R_s	κ	κ_w
	1	2	3	4	5	6	7			
Case 1	▢	▢	▢	▢	▢	▢	▢	0.675	0.265	0.328
Case 2	▢	▢	▢	▢	▢	▢	▢	0.704	0.282	0.338
Case 3	▢	▢		▢	▢	▢	▢	0.845	0.431	0.559
Case 4	▢	▢		▢	▢			0.785	0.509	0.581
Case 5	▢	▢				▢	▢	0.709	0.382	0.604
Case 6	▢	▢		▢	▢	▢		0.848	0.848	0.848
Case 7	▢	▢		▢		▢	▢	0.814	0.716	0.764

off-diagonal values are expected to represent the similarities between the corresponding classes. The columnar class is assumed to resemble none of the other classes. We compute the statistics R_s, κ, and κ_w for seven different experiments with the following settings:

- *Case*1: We order all of the cells in the whole data using the optimal leaf ordering algorithm by maximizing the sum of similarities between the adjacent leaves as in (10).
- *Case*2: We order all of the cells in the whole data using the optimal leaf ordering algorithm by maximizing the sum of

similarities between every leaf and the leaves in its adjacent clusters as in (11). Hereafter, we will use this criterion for the optimal leaf ordering algorithm because it provided better results as shown in Table 6.

- *Case*3: We order the cells of all classes except the columnar cells. The columnar cells are rarely encountered in the images
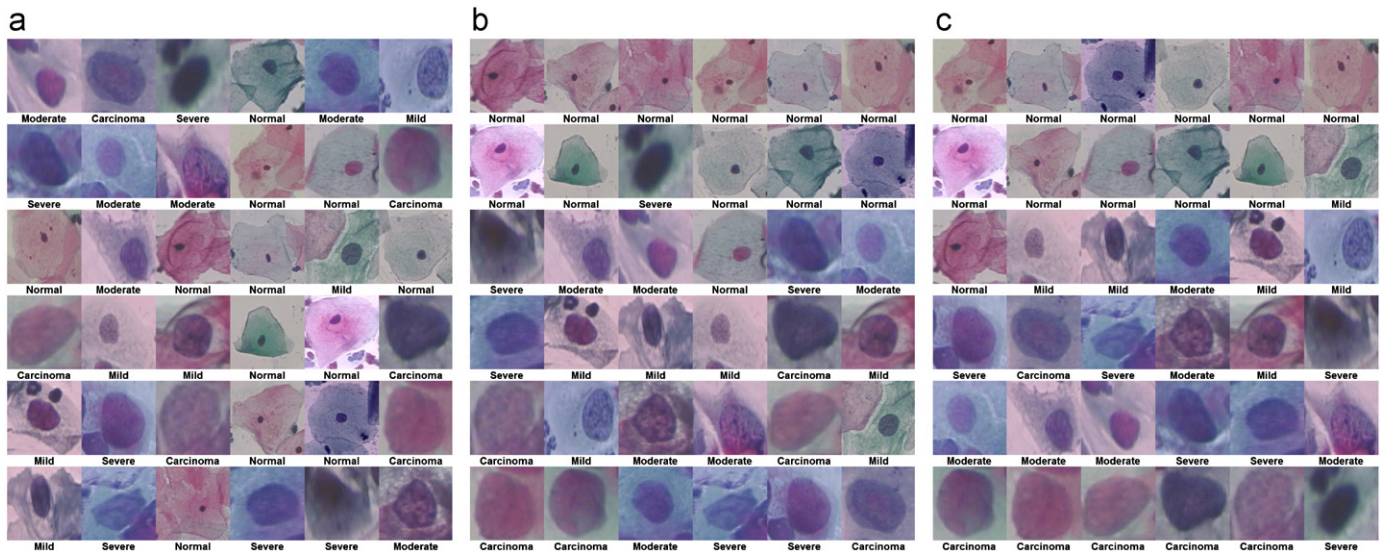
Fig. 17. Comparison of different orderings. (a) Random arrangement: $R_s=0.038$, $\kappa=-0.100$, and $\kappa_w=-0.054$. (b) Initial ordering resulting from hierarchical clustering: $R_s=0.771$, $\kappa=0.266$, and $\kappa_w=0.417$. (c) Result of optimal ordering: $R_s=0.895$, $\kappa=0.466$, and $\kappa_w=0.614$. The images are resized to the same width and height so the relative sizes of the cells are not proper.

of the Hacettepe data, and we do not include the columnar cells in the rest of the experiments.

- *Case4*: We order the cells of all classes except the columnar, severe dysplasia, and carcinoma in situ classes. In this way, we aim to evaluate the performance when we are given Pap test images of patients at early stages of the disease.
- *Case5*: We order the cells of all classes except the columnar, mild dysplasia, and moderate dysplasia classes. In this way, we aim to evaluate the performance when we are given Pap test images of patients at late stages of the disease.
- *Case6*: First, we group the superficial squamous and intermediate squamous classes into a single class called normal, and the mild dysplasia, moderate dysplasia, severe dysplasia, and carcinoma in situ classes into a single class called abnormal. Then, we evaluate the performance using two classes, namely normal and abnormal.
- *Case7*: We again group the superficial squamous and intermediate squamous classes into a single class called normal. Next, the mild dysplasia and moderate dysplasia classes are grouped into a single class called early-abnormal, and the severe dysplasia and carcinoma in situ classes are grouped into a single class called abnormal. Then, we evaluate the performance using three classes, namely normal, early-abnormal, and abnormal.

Table 6 summarizes the experimental results obtained for different settings. The performance improved when there were no columnar cells in the input data. Dropping columnar cells does not lead to an unrealistic situation, because the Pap test images of the Hacettepe data set rarely included columnar cells. We obtained an almost perfect agreement by grouping the data into normal and abnormal classes for which both kappa coefficients κ and κ_w were calculated as greater than 0.8. This supports the conjecture that the cervical cells can be grouped according to their abnormality degree using our ranking method. Moreover, we achieved a substantial agreement when we grouped the mild dysplasia and moderate dysplasia as well as the severe dysplasia and carcinoma in situ classes (cases 6 and 7). This setting is appropriate because the severe dysplasia and carcinoma in situ classes are considered as very similar [39]. We also performed the same experiments using only the nucleus features (9 out of 14). The results are not included here due to space constraints but the performance of using all features was significantly better for

all settings. This shows the importance of cytoplasm features even though the segmentation of cytoplasm regions remained approximate in our segmentation algorithm whose main focus was to accurately delineate the nuclei regions. Further improvements in cytoplasm segmentation can improve the overall accuracy in future work.

Fig. 17 shows three different orderings for an example set of cells. We can observe that the second ordering that corresponds to the initial ordering resulting from hierarchical clustering is superior to the first one that was a randomly generated arrangement, and the third ordering that corresponds to the result of the optimal leaf ordering algorithm is superior to the second one by visual inspection. This observation is consistent with the information provided by the statistical coefficients that we use to measure the agreement between the result of our ranking procedure and the ground truth. Indeed, all three coefficients indicate the best agreement for the third ordering and the worst agreement for the first ordering. Moreover, the kappa coefficients are less than 0 for the first ordering, meaning that the agreement is actually worse than chance.

6.4. Computational complexity

The proposed algorithms were implemented in Matlab. The overall processing using the unoptimized Matlab code took 261 seconds (s) on the average for 2048×2048 pixel Hacettepe images on a PC with a 3.4 GHz Intel Core i7 processor and 4 GB RAM. The running times were obtained on images having an average number of 45 cells. We performed a code profile analysis to investigate the time spent in different steps. Among the major steps, on the average, background extraction took 20.8 s (including 2.4 s for RGB to Lab color transformation, 4.7 s for illumination correction, and 13.7 s for thresholding), segmentation of cells took 235 s (including 194 s for region hierarchy construction and 41 s for region selection and classification), and classification of cells took 4.9 s (including 4.5 s for feature extraction and 0.42 or 0.39 s for ranking using adjacent leaves or adjacent clusters, respectively). These times can be significantly reduced by optimizing the Matlab code or by implementing some of the steps in C if needed.

7. Conclusions

We presented a computer-assisted screening procedure that aimed to help the grading of Pap smear slides by sorting the cells according to their abnormality degrees. The procedure consisted of two main tasks. The first task, segmentation, involved morphological operations and automatic thresholding for isolating the cell regions from the background, a hierarchical segmentation algorithm that used homogeneity and circularity criteria for partitioning these cells, and a binary classifier for separating the nuclei from cytoplasm within the cell regions. The second task, classification, ranked the cells based on their feature characteristics computed from the nuclei and cytoplasm regions. The ranking was generated via linearization of the leaves of a binary tree that was constructed using hierarchical clustering. Experiments using two data sets showed that the proposed approach could produce accurate segmentation and classification of cervical cells in images having inconsistent staining, poor contrast, and overlapping cells. Furthermore, both the segmentation and the classification algorithms are parameter-free and generic so that additional criteria can easily be incorporated to improve the identification of different cell types without any requirement for training examples.

Our main focus in this paper was to correctly delineate the individual nuclei while assuming that the overlapping cytoplasm areas were shared by different cells. We observed that, in addition to nuclei features, using the cytoplasm features extracted from this approximate segmentation resulted in an increase in the classification accuracy. Therefore, future work will include additional steps for improving the segmentation of nuclei and cytoplasm areas within overlapping cell groups. We believe that more accurate features computed from the improved cytoplasm regions will also improve the grading performance.

Acknowledgment

References

[1] World Health Organization, Comprehensive Cervical Cancer Control: A Guide to Essential Practice, 2006.

[2] H.Z. Noorani, C. Arratoon, A. Hall, Assessment of Techniques for Cervical Cancer Screening, Technical Report CCOHTA Report 1997: 2E, Canadian Coordinating Office for Health Technology Assessment (May 1997).

[3] E. Bengtsson, Computerized cell image analysis: past, present, and future, in: Proceedings of 13th Scandinavian Conference on Image Analysis, 2003, pp. 395–407.

[4] E. Bengtsson, Recognizing signs of malignancy—the quest for computer assisted cancer screening and diagnosis systems, in: IEEE International Conference on Computational Intelligence and Computing Research, 2010.

[5] P. Bamford, B. Lovell, Unsupervised cell nucleus segmentation with active contours, Signal Processing 71 (2) (1998) 203–213.

[6] H.-S. Wu, J. Barba, J. Gil, A parametric fitting algorithm for segmentation of cell images, IEEE Transactions on Biomedical Engineering 45 (3) (1998) 400–408.

[7] S.-F. Yang-Mao, Y.-K. Chan, Y.-P. Chu, Edge enhancement nucleus and cytoplasm contour detector of cervical smear images, IEEE Transactions on Systems, Man, and Cybernetics Part B: Cybernetics 38 (2) (2008) 353–366.

[8] M.-H. Tsai, Y.-K. Chan, Z.-Z. Lin, S.-F. Yang-Mao, P.-C. Huang, Nucleus and cytoplasm contour detector of cervical smear image, Pattern Recognition Letters 29 (9) (2008) 1441–1453.

[9] I. Dagher, K.E. Tom, WaterBalloons: a hybrid watershed balloon snake segmentation, Image and Vision Computing 26 (7) (2008) 905–912.

[10] P.-W. Huang, Y.-H. Lai, Effective segmentation and classification for HCC biopsy images, Pattern Recognition 43 (4) (2010) 1550–1563.

[11] N.M. Harandi, S. Sadri, N.A. Moghaddam, R. Amirfattahi, An automated method for segmentation of epithelial cervical cells in images of ThinPrep, Journal of Medical Systems 34 (6) (2010) 1043–1058.

[12] M.E. Plissiti, C. Nikou, A. Charchanti, Automated detection of cell nuclei in Pap smear images using morphological reconstruction and clustering, IEEE Transactions on Information Technology in Biomedicine 15 (2) (2011) 233–241.

[13] M.E. Plissiti, C. Nikou, A. Charchanti, Combining shape, texture and intensity features for cell nuclei extraction in Pap smear images, Pattern Recognition Letters 32 (6) (2011) 838–853.

[14] K. Li, Z. Lu, W. Liu, J. Yin, Cytoplasm and nucleus segmentation in cervical smear images using Radiating GVF snake, Pattern Recognition 45 (4) (2012) 1255–1264.

[15] Y. Marinakis, G. Dounias, J. Jantzen, Pap smear diagnosis using a hybrid intelligent scheme focusing on genetic algorithm based feature selection and nearest neighbor classification, Computers in Biology and Medicine 39 (1) (2009) 69–78.

[16] A. Kale, S. Aksoy, S. Onder, Cell nuclei segmentation in pap smear test images, in: Proceedings of 17th IEEE Signal Processing and Communications Applications, Antalya, Turkey, 2009.

[17] S. Ali, A. Madabhushi, Active contour for overlap resolution using watershed based initialization (ACOReW): applications to histopathology, in: IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2011, pp. 614–617.

[18] I. Vanhamel, I. Pratikakis, H. Sahli, Multiscale gradient watersheds of color images, IEEE Transactions on Image Processing 12 (6) (2003) 617–626.

[19] Y.-Y. Wang, Y.-N. Sun, C.-C. K. Lin, M.-S. Ju, Nerve cell segmentation via multi-scale gradient watershed hierarchies, in: 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2006, pp. 4310–4313.

[20] M.A.G. de Carvalho, R.deA. Lotufo, M. Couprie, Morphological segmentation of yeast by image analysis, Image and Vision Computing 25 (1) (2007) 34–39.

[21] R.F. Walker, P. Jackway, B. Lovell, I. D. Longstaff, Classification of cervical cell nuclei using morphological segmentation and textural feature extraction, in: Proceedings of 1994 Second Australian and New Zealand Conference on Intelligent Information Systems, 1994, pp. 297–301.

[22] Y.-Y. Chou, L.G. Shapiro, A hierarchical multiple classifier learning algorithm, Pattern Analysis and Applications 6 (2) (2003) 150–168.

[23] J. Zhang, Y. Liu, Cervical cancer detection using svm based feature screening, in: International Conference on Medical Image Computing and Computer Aided Intervention, vol. 2, 2004, pp. 873–880.

[24] H.C. Kitchener, R. Blanks, G. Dunn, L. Gunn, M. Desai, R. Albrow, J. Mather, D.N. Rana, H. Cubie, C. Moore, R. Legood, A. Gray, S. Moss, Automation-assisted versus manual reading of cervical cytology (MAVARIC): a randomised controlled trial, Oncology 12 (1) (2011) 56–64.

[25] E. Martin, Pap-smear Classification, Master's Thesis, Technical University of Denmark, Oersted-DTU, Automation, Lyngby, Denmark, 2003.

[26] P. Soille, Morphological Image Analysis, second ed., Springer, 2002.

[27] J. Kittler, J. Illingworth, Minimum error thresholding, Pattern Recognition 19 (1) (1986) 41–47.

[28] A. Kale, S. Aksoy, Segmentation of cervical cell images, in: Proceedings of 20th IAPR International Conference on Pattern Recognition, Istanbul, Turkey, 2010, pp. 2399–2402.

[29] E.R. Dougherty, R.A. Lotufo, Hands-on Morphological Image Processing, SPIE Publications, 2003.

[30] R.M. Haralick, L.G. Shapiro, Computer and Robot Vision, Addison-Wesley, 1992.

[31] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, Numerical Recipes: The Art of Scientific Computing, third ed., Cambridge University Press, 2007.

[32] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, John Wiley & Sons, Inc., 2000.

[33] H.G. Akcay, S. Aksoy, Automatic detection of geospatial objects using multiple hierarchical segmentations, IEEE Transactions on Geoscience and Remote Sensing 46 (7) (2008) 2097–2111.

[34] Z. Bar-Joseph, D.K. Gifford, T.S. Jaakkola, Fast optimal leaf ordering for hierarchical clustering, Bioinformatics 17 (2001) S22–S29.

[35] B. Ozdemir, S. Aksoy, S. Eckert, M. Pesaresi, D. Ehrlich, Performance measures for object detection evaluation, Pattern Recognition Letters 31 (10) (2010) 1128–1137.

[36] A. Hoover, G. Jean-Baptiste, X. Jiang, P.J. Flynn, H. Bunke, D.B. Goldgof, K. Bowyer, D.W. Eggert, A. Fitzgibbon, R.B. Fisher, An experimental comparison of range image segmentation algorithms, IEEE Transactions on Pattern Analysis and Machine Intelligence 18 (7) (1996) 673–689.

[37] A.P. Zijdenbos, B.M. Dawant, R.A. Margolin, A.C. Palmer, Morphometric analysis of white matter lesions in MR images: method and validation, IEEE Transactions on Medical Imaging 13 (4) (1994) 716–724.

[38] R.W. Mackin Jr., B. Roysam, J.N. Turner, Adaptive 3-D segmentation algorithms for microscope images using local in-focus, and contrast features: application to Pap smears, in: IEEE International Conference on Image Processing, vol. 3, 1995, pp. 160–163.

[39] J. Norup, Classification of Pap-smear Data by Transductive Neuro-fuzzy Methods, Master's Thesis, Technical University of Denmark, Oersted-DTU, Automation, Lyngby, Denmark, 2005.

Aslı Gençtav received the B.S. degree from the Middle East Technical University (METU), Ankara, Turkey, in 2007. She received the M.S. degree from the Bilkent University, Ankara, Turkey, in 2010. She is currently a Ph.D. candidate at the Department of Computer Engineering at METU. Her research interests include computer vision.

Selim Aksoy received the B.S. degree from the Middle East Technical University, Ankara, Turkey, in 1996 and the M.S. and Ph.D. degrees from the University of Washington, Seattle, in 1998 and 2001, respectively. Since 2004, he has been an Assistant Professor at the Department of Computer Engineering, Bilkent University, Ankara. His research interests include computer vision, statistical and structural pattern recognition with applications to remote sensing, medical imaging, and multimedia data analysis. He is an Associate Editor of Pattern Recognition Letters.

Sevgen Önder finished his pathology residency at the Hacettepe University School of Medicine in 2003. Currently he works in the same department as an Assistant Professor. His major interests are cytopathology and pulmonary pathology. In addition to morphology, he is also interested in molecular pathology and utilization of advanced technology for diagnostic purpose.