Cytometry
PART A
Journal of the
International Society for
Advancement of Cytometry

Automatic Segmentation of Cell Nuclei in Feulgen-Stained Histological Sections of Prostate Cancer and Quantitative Evaluation of Segmentation Results

Birgitte Nielsen,[1,2] Fritz Albregtsen,[1,3] Håvard E. Danielsen[1,2,3]*

[1]Institute for Medical Informatics, Oslo University Hospital, Norway

[2]Centre for Cancer Biomedicine, University of Oslo, Norway

[3]Department of Informatics, University of Oslo, Norway

*Correspondence to: Håvard E. Danielsen, Institute for Medical Informatics, Radiumhospitalet, Oslo University Hospital, Montebello, N-0310 Oslo, Norway

Email: havard.danielsen@ oslo-universitetssykehus.no

ISAC
International Society for Advancement of Cytometry

● **Abstract**
Digital image analysis of cell nuclei is useful to obtain quantitative information for the diagnosis and prognosis of cancer. However, the lack of a reliable automatic nuclear segmentation is a limiting factor for high-throughput nuclear image analysis. We have developed a method for automatic segmentation of nuclei in Feulgen-stained histological sections of prostate cancer. A local adaptive thresholding with an object perimeter gradient verification step detected the nuclei and was combined with an active contour model that featured an optimized initialization and worked within a restricted region to improve convergence of the segmentation of each nucleus. The method was tested on 30 randomly selected image frames from three cases, comparing the results from the automatic algorithm to a manual delineation of 924 nuclei. The automatic method segmented a few more nuclei compared to the manual method, and about 73% of the manually segmented nuclei were also segmented by the automatic method. For each nucleus segmented both manually and automatically, the accuracy (i.e., agreement with manual delineation) was estimated. The mean segmentation sensitivity/specificity were 95%/96%. The results from the automatic method were not significantly different from the ground truth provided by manual segmentation. This opens the possibility for large-scale nuclear analysis based on automatic segmentation of nuclei in Feulgen-stained histological sections.   © 2012 International Society for Advancement of Cytometry

DIGITAL pathology is one of the main application fields of automated medical image analysis. A steadily improving imaging technology has given us an increasing volume of high quality medical images. This increasing volume of images, both in routine clinical work and in research and development, calls for an increasing degree of automation of image analysis. We are also at the brink of moving away from subjective evaluation being done at the microscope, toward automated and more objective diagnostic and prognostic work, provided that we can ensure the quality of quantitative measurements of relevant image features.

Digital image analysis of cell nuclei is a very useful method to obtain quantitative information for the diagnosis and prognosis of human cancer (1). But before we can extract meaningful parameters describing the cell nuclei of a specimen, we have to segment the nuclei from the rest of the image. Nuclei in tissue sections are in general difficult to segment by purely automatic means because; (1) the cells may be clustered, (2) the image background varies, (3) there are intensity variations within the nuclei, and (4) the nuclear boundary may be diffuse, either along the whole perimeter or in sectors of varying width. Fluorescence microscopy images of nuclei in

tissue sections often show an uneven background, due to autofluorescence from the tissue and fluorescence from objects that are out of focus, while light-microscopy images of nuclei in tissue sections are even more complex due to the presence of a visible background (2). Therefore, segmentation of nuclei in light-microscopy images of thin sections is often done in a semiautomatic or interactive way (3–6). The lack of a reliable automatic nuclear segmentation is a limiting factor for high-throughput nuclear image analysis in light-microscopy images of routine histopathological sections.

A large number of methods for segmentation of cells and cell nuclei for several biomedical applications have been published. A large amount of recently published segmentation methods are applied on fluorescence images (7–20), and examples of applications on such images are spatial analysis of DNA sequences and nuclear structures, nuclei tracking and studies of surface-stained living cells. In our institute, we have developed a system for high-throughput nuclear texture analysis to perform a prognostic classification of cancer patients, based on up to 50,000 measured nuclei/case. Using light-microscopy images of Feulgen-stained sections we obtain more detailed morphological information compared to fluorescence images.

Segmentation of cell nuclei can be viewed as an object modeling problem. Successful global thresholding requires that the nuclei have a range of intensities that is sufficiently different from the background. This is generally not true, since the background varies. The result may be improved by adaptive thresholding, but large intensity variations between and within the nuclei will cause the model to fail. Region growing methods, e.g., the watershed algorithm, are based on the assumption that the objects consist of connected regions of similar pixels. Region growing methods combined with region-merging are commonly used in segmentation of cells and nuclei from fluorescence images (7,10). Again, large intensity variations between and within the nuclei may cause the model to fail. Edge-based segmentation models the nucleus as an entity bounded by an intensity gradient. In practice, the detected edges are not always sharp, they do not cover the whole circumference of all nuclei, and clustered cells pose a problem. Active contour models or snakes (21) are widely used in medical image segmentation (11,12,15,16,22,23). However, these methods are sensitive to initialization of a start contour or a seed inside each nucleus. Baggett et al. (18) and McCullough et al. (19) proposed semiautomatic methods based on dynamic programming for segmentation of cells and nuclei, requiring the user to mark two points per cell (or nucleus), one approximately in the center and the other on the border.

Bengtsson et al. (2) discussed the relative advantages of different approaches to cell segmentation, and concluded that more robust segmentation can be obtained if a combination of cellular features such as intensity, gradient, and shape is used. Based on this, Bengtsson et al. (2) and Wählby et al. (7) proposed a seeded watershed algorithm as the most useful tool for incorporating such features into the cell model. They presented a seeded watershed segmentation of the gradient

magnitude image in which seeds representing both object and background are created by combining morphological filtering of both the original image and the gradient magnitude image. Subsequent steps (merging of weak borders and cluster separation) refined and improved the results. Plissiti et al. (23) proposed a seeded watershed segmentation of nuclei in conventional Pap stained cervical smear images that combined intensity, gradient, and shape information. They found that the proposed method produced more accurate nuclear segmentation compared to a gradient vector flow deformable model and a region based active contour model. Malpica et al. (20) proposed a watershed algorithm combining gradient and shape information for splitting clusters of nuclei in blood and bone marrow preparations.

The aim of the present study has been to develop a method for automatic segmentation of cell nuclei in light-microscopy images of Feulgen-stained histological sections of prostate cancer. We have not found any recent studies on automatic segmentation of nuclei from such prostate sections. We have followed the combined approach of Bengtsson et al. (2), but we have combined adaptive thresholding with an active shape model. Initial tests using global thresholding, edge detection, watershed with different region merging techniques, and also the method of Wahlby et al. (2,7) gave poor results on our images. But it is difficult to compare our approach with the method of Wahlby et al., because the material, the staining, and the imaging modalities are so different.

Our combined approach for automatic segmentation of nuclei from Feulgen-stained sections of prostate cancer is based on a number of interacting steps from several strategies traditionally used in image segmentation: A gradient-validated local adaptive thresholding and an active contour model that features an optimized initialization and works within a restricted annular region to improve convergence of the segmentation of each nucleus. We propose a method based on several steps to; (1) detect the nuclei, (2) optimize the initial contours for the snakes by a coarse segmentation, (3) optimize the convergence of the snakes, and (4) split overlapping segmentation masks.

To detect the nuclei and make binary "marker" images to find a start contour for each snake we use local adaptive thresholding. On the basis of the evaluation of Trier and Taxt (24) and Trier and Jain (25) we have used the Niblack method (26) with the postprocessing step of Yanowitz and Bruckstein (27). Initial tests using a traditional snake (21) required a very precise initialization, and therefore gave poor results. In the present study, we therefore used the gradient vector flow (GVF) snake (28,29).

The evaluation of the segmentation result is a very important issue, and a quantitative evaluation is vitally important, as image data sets should be much too large to oversee manually. Udupa et al. (30,31) considered three factors when evaluating segmentation methods: precision (reproducibility in the presence of data variability), accuracy (agreement with the real object), and efficiency (computer and user time needed). In the evaluation of medical image segmentation results, we cannot always rely on the manual segmentation as the "ground

truth," as it is affected by intra- and interexpert variability. In addition, the merging of "ground truths" from several experts into "latent gold standards" is not trivial (32,33). A number of validation frameworks and metrics have been proposed, but there is no general consensus on which approach to use. Metrics from general computer vision that treat under- and oversegmentation as equivalent errors of accuracy will run into trouble in some medical applications (31). So segmentation accuracy has to be combined with knowledge about clinical relevancy and the medical impact of different errors (34).

A very small number of samples seem to be a common characteristic of many papers on evaluation of accuracy in medical image segmentation. Popovic et al. (34) suggest a new validation metric to assess segmentation accuracy, but includes only a single CT image of calvarial tumors in six different patients as a case study. Udupa et al. (30) outlines the methodology for evaluation, but applies it to a single MR image of a patient's brain. Warfield (33) proposed a method to estimate the accuracy of automated segmentation as compared to a group of expert segmentations, and simultaneously measure the quality of each expert, but illustrated its application on just three different MRI images. Einstein et al. (35) discuss and compare reproducibility (agreement among repeated measurements) and accuracy (agreement between measurement and external standard) in interactive cell segmentation. Cytological material from six patients with invasive ductal carcinoma of the breast was used, but only six nuclei per patient were segmented. The method of Wählby et al. (7) was tested on six different 2D images containing a total of 689 cells, and gave about 90% correct segmentation, in the sense that it tallied with manual counts from the same image fields. Thus, the geometrical accuracy of the segmentation of each cell was not considered.

Our segmentation method was developed using a "training set" of two frame images taken from one case of prostate cancer and was evaluated using 30 randomly selected frames (including the two training frames) from three cases. The evaluation was performed by comparing the segmentation results with a manual delineation of 924 nuclei from the 30 frames.

Materials and Methods

Materials

Patients. The use of clinical material was approved by the Norwegian Regional Committees for Medical Research Ethics (REK). Tissue samples from three men with early prostate cancer who underwent radical prostatectomy were included in the study.

Tissue preparation. Paraffin embedded tissue samples fixed in 4% buffered formalin were cut into 5-μm sections. The tissue sections were first stained with hematoxylin and eosin (H&E). After removing the coverslip with xylol, the sections were rehydrated and then de-stained for 1 h in 1% HCl, followed by fixation and restaining with Feulgen. A pathologist marked out the tumor area on an image of the H&E stained

section. The coordinates of the area where then transferred to a high-resolution automatic imaging system to allow acquisition of images from the correct area of the Feulgen-stained section.

Image Acquisition and Preprocessing

The Zeiss AxioImager.Z1 automated microscope equipped with a 63×/1.4 objective lens, a 546 nm green filter and a black and white high-resolution digital camera (Axiocam MRM, Zeiss) with 1,040 × 1,388 pixels and a gray level resolution of 12 bits per pixel was used to capture each frame image. The pixel resolution was 102.5 nm per pixel on the specimen.

Shade correction. Varying illumination, reflectance, and optical transmission over the field of view may play a role in the success of any image segmentation algorithm based on gray level thresholding. Although a local adaptive thresholding method is used to detect the cell nuclei within each frame image, we still want to remove such shading effects, dividing each frame image by a background image. Before capturing frame images of a given case, a background image was captured on a blank area of the slide outside the specimen (see Supporting Information).

Noise removal and frame image resizing. A 5 × 5 median-filter was applied on the shade corrected frame images in the object detection step of the segmentation method. This was done to reduce possible noise without too much unwanted altering of local texture. For the object detection step, we do not need the full resolution of the frame images. To speed up the processing, the median-filtered frame images were therefore resized by a scale factor of 0.5 (i.e., shrunk by a factor of two) using bicubic interpolation and antialiasing, i.e., the output pixel value was a weighted average of the pixels in the nearest 4 × 4 neighborhood (36).

The Segmentation Method

The proposed segmentation method is based on a number of interacting (sometimes iterative) steps that combine several strategies traditionally used in image segmentation (see Fig. 1). Following a background correction and noise reduction, a local adaptive thresholding detected the nuclei. The objects were filtered by morphological operators, and small objects were removed. The resulting binary markers were used to extract subimages of individual nuclei (or clusters of nuclei). Based on the convex hull solidity of each marker (i.e., the proportion of the pixels in the convex hull that are also in the marker), clusters were split by iterative erosion and dilation. Object masks were reproduced from the markers by dilation with the same structuring elements as were used for producing the markers.

Dilation and erosion were then used to produce a start contour and a bounded, annular edge map for a GVF snake to operate on, producing the final segmentation of each nucleus. Segmentation masks (and reproduced object masks) for nuclei
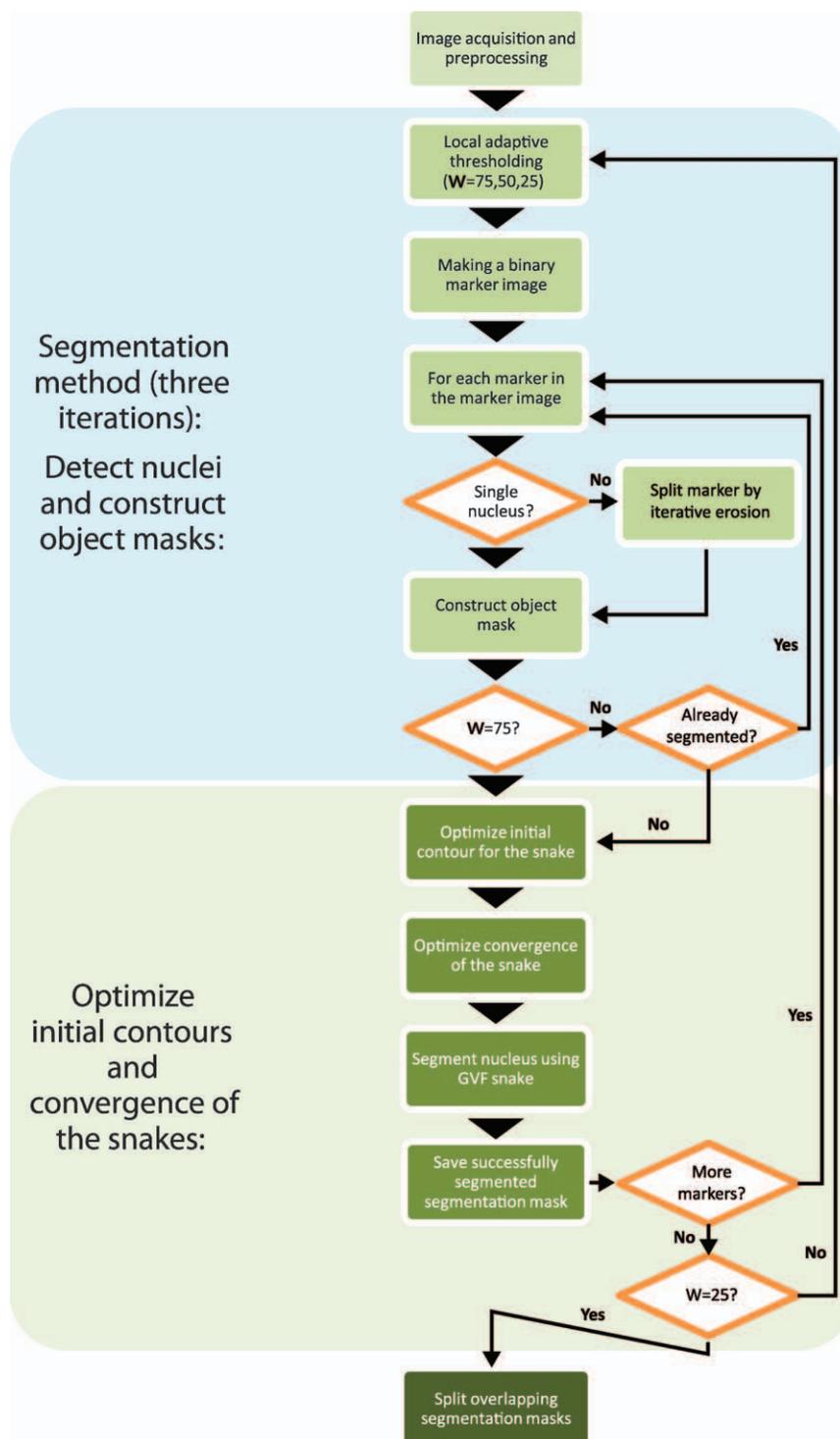
Figure 1. Flow diagram illustrating the proposed segmentation method. The nuclei were detected by local adaptive thresholding. For each pixel in the frame image a gray level threshold was computed based on the gray level values within a $W \times W$ pixels window. Because both the size of the nuclei and the distance between the nuclei varies, the method was iterated three times corresponding to three different window sizes in the local adaptive thresholding ($W = 25,50,75$). Binary markers and corresponding object masks representing single nuclei were produced using erosion and dilation. Each object mask was then used to produce a start contour and a bounded, annular edge map for a GVF snake to operate on, producing the final segmentation of each nucleus.

that were considered successfully segmented in the first iteration of the method (corresponding to a window size of $W \times W = 75 \times 75$ in the thresholding) were stored in binary ("reproduced object mask" and "segmentation mask") images. By segmentation mask we mean the mask that was created for each segmented nucleus after the final segmentation of each nucleus using the GVF snake. By object mask we mean the mask that was reproduced from a marker that was obtained by thresholding and erosion.

For each of the following two iterations (window size 50×50 and 25×25, see Fig. 1), the reproduced object masks (produced in these iterations) were compared to the "reproduced object mask" image to decide whether the nuclei were already segmented in a previous iteration. Nuclei that were not segmented in a previous iteration were further processes by GVF snakes.

The result of the proposed method depends on a number of parameter values that were specified by trial and error on the two training frames. A discussion of the robustness of the method to changes in parameter values is included in the Result section.

The segmentation method was developed in Matlab (36). In addition, the method has been implemented in C++ on a computer cluster (consisting of Intel Xeon 2.5 GHz processors) at our institution. The Matlab version of the software can be accessed from the link: http://medinfo.net/segmentationfiles/. We used Matlab 7.5.0 (R2007b).

Snakes. The idea behind active contours for image segmentation is quite simple; (1) the user specifies an initial guess for the contour, and (2) the contour is moved by image driven forces to the boundary of the object, while being restrained by internal forces stemming from the curve itself. The traditional snake (21) is an energy minimizing parametric contour $\vec{x}(s) = [x(s), y(s)]$, $s \in [0, 1]$. The final position of the snake corresponds to a minimum of the energy functional

$$E(\vec{x}) = \int_0^1 \frac{1}{2}\left[\alpha\left|\frac{\partial \vec{x}}{\partial s}\right|^2 + \beta\left|\frac{\partial^2 \vec{x}}{\partial s^2}\right|^2\right] + P(\vec{x}) \; ds \qquad (1)$$

where α and β are weighting parameters that control the internal tension and rigidity of the snake, respectively. The external energy is given by an integration of a scalar potential function defined over the image, along the parametric curve that defines the snake.

The external energy functional

$$P(\vec{x}) = -\left|\nabla\left[G_\sigma(x, y) * f(x, y)\right]\right|^2 \qquad (2)$$

where $G_\sigma(x,y)$ is a 2D Gaussian function with standard deviation σ, ∇ is the gradient operator, and $f(x,y)$ is the image, will attract the snake to locations with large image gradients, e.g., lines, edges and object contours in the image.

A general problem is that the initial contour needs to be relatively close to the target object boundary to get convergence. This is known as the "capture range problem." Another problem is poor convergence to boundary concavities. Relatively large σ-values may be needed to increase the capture range, but this will also substantially blur the object boundaries that we want the curve to converge to.

Xu and Prince (28) addressed these problems by using the gradient vector flow (GVF), computed as a diffusion of the gradient vectors of a gray level or binary edge map derived from the image. The GVF is defined as the vector field $\vec{v}(x,y) = [u(x,y), v(x,y)]$ that minimizes the following energy functional

$$\varepsilon = \int\int \mu(u_x^2 + u_y^2 + v_x^2 + v_y^2) + |\nabla e|^2|\vec{v} - \nabla e|^2 dxdy \qquad (3)$$

where u_x, u_y, v_x, and v_y are the partial derivatives of the vector field and ∇e is the gradient of an edge map (e can be any gray level or binary edge map defined in the image processing literature). For high values of ∇e, $\vec{v}$ will be nearly equal to ∇e, while for low values of ∇e (corresponding to homogenous regions), $\vec{v}$ will vary slowly. The parameter μ is governing the tradeoff between the two terms in the integrand and should be set according to the amount of noise in the image. In the present study, we have used the GVF toolbox (29).

Local adaptive thresholding. Although a shade correction was performed, a global thresholding did not produce satisfactory results. Therefore, local adaptive thresholding had to be considered. Trier and Taxt (24) and Trier and Jain (25) evaluated 11 methods on grayscale document images with low contrast, variable background intensity and noise. Among these, the method of Yanowitz and Bruckstein (27) includes an object perimeter gradient computation step to remove false objects in the segmented image. As pointed out by Yanowitz and Bruckstein and demonstrated in the surveys (24,25), incorporating this validation step will improve other segmentation methods as well. Based on the finding that the Niblack method (26) with this gradient-validation step incorporated performed the best and was also one of the fastest methods, we have used this scheme as a part of our segmentation method.

Thus, for each pixel (x,y) in the frame image $f(x,y)$ a local threshold was calculated as $t(x,y) = m(x,y) + \delta \cdot s(x,y)$, where m and s are the mean and standard deviation of the gray level values within $W \times W$ pixels window of (x,y) and δ is a constant. We found that a δ-value of 0.2 gave good results. The size of the window should be related to the size of the nuclei. The size of the minimum bounding box of typical nuclei (i.e., the smallest rectangle containing the nucleus) in our resized training frame images varied from 15×15 to 50×50 pixels. Generally, the window size should be larger than the nuclei (up to 50% larger than the minimum bounding box) to produce good object masks covering the complete nuclei. We found that a window size of 75×75 pixels gave good results on the largest nuclei in the resized frame images. This window size also gave good results on smaller nuclei provided that they were well separated from other nuclei. However, smaller nuclei close to each other were better separated by using smaller window sizes. The adaptive thresholding results were very robust to small changes in window size. Therefore, we used only three different values, i.e., $W = 75$, 50, and 25, corre-
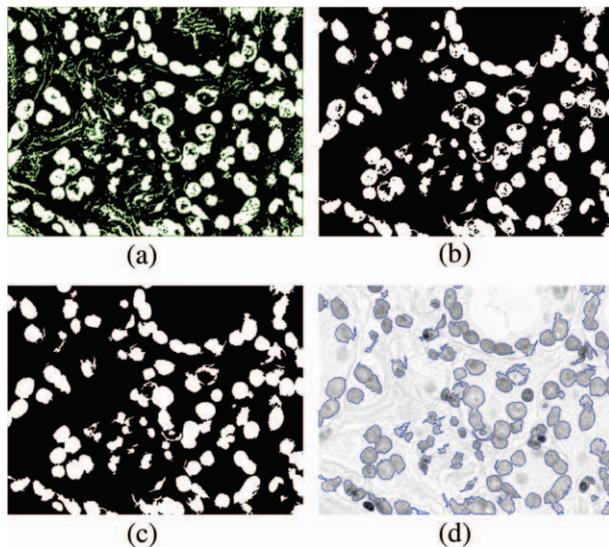
Figure 2. (**a**) The binary frame image obtained after local adaptive thresholding of the preprocessed, resized gray scale frame image. The boundaries of all the connected components are shown in green. (**b**) The result after the gradient validation step of Yanowitz and Bruckstein. (**c**) The result after filling holes and removing small objects. (**d**) The boundaries of the resulting objects plotted as blue boundaries on the frame image. A window size of 75×75 pixels was used in the local adaptive thresholding.

sponding to three iterations of the segmentation method (Fig. 1). Figure 2a shows a binary frame image obtained using the local threshold $t(x,y)$ computed in the first iteration of the segmentation method.

The postprocessing step of Yanowitz and Bruckstein (27) was then applied on the thresholded frame image: First, all the connected components (objects and holes) in the thresholded image were labeled. This was done by computing a label matrix L, i.e., a 2D matrix of non-negative integers that represent contiguous regions. The k-th region includes all elements in L that have value k. The zero-value elements of L make up the background (36). Then, the gradient magnitude of the gray level frame image was computed by convolving with the derivative of a Gaussian. The standard deviation of the Gaussian filter was set to $\sigma = 2$. Finally, the boundaries of all labeled objects and holes (Fig. 2a) were traced and used to compute the average gradient magnitude of each object (or hole) boundary in the gray level image. If the average gradient magnitude of a given boundary did not exceed a certain threshold, then the object was regarded as not being a nucleus and therefore removed from the image. We used a gradient threshold value of 0.1. Using this threshold value, we found that false objects were removed, while true nuclei were kept (Fig. 2b).

Making binary marker images. In our thin section material, we have observed holes in the object masks that are representing interior parts of the nuclei containing a small amount of Feulgen stain. We have not observed holes representing a gap between tightly clustered nuclei. The holes in the binary frame image were filled, using an algorithm based on morphological

reconstruction (36,37). A hole was defined as a set of background pixels that cannot be reached by filling in the background from the edge of the image (36). The 4-connected background neighbors were used. We obtained good object masks covering the complete nuclei for most of the in-focus, nonoverlapping complete nuclei. Object masks representing only a part of the nuclei (e.g., C-shaped object masks, see Fig. 2) may occur for nuclei containing a very small amount of Feulgen stain both in the interior part and along parts of the nuclear boundary.

Finally, small objects (area $\leq$ 10 pixels) were removed from the image. The resulting binary image containing objects with area $>$ 10 pixels are shown in Figure 2c. The frame image with the resulting object boundaries is shown in Figure 2d.

The binary frame images were then resized back to their original size. To produce "markers" for the nuclei or for clusters of two or more nuclei (a marker is defined here as a subset of an object mask), the binary image was eroded with a disk shaped structuring element with a radius $r_s = 15$ pixels. The size of the structuring element was chosen to: (i) produce a large number of markers representing single nuclei and (ii) avoid deleting small object masks (corresponding to small nuclei). In addition, some markers representing two or more nuclei were obtained. An object mask representing only a part of a nucleus, e.g., a C-shaped mask, may result in one or more erroneous markers representing only parts of the nucleus. However, erroneous segmentations resulting from such erroneous markers are handled either by the segmentation algorithm (by checking some properties of the segmentation masks, see saving the segmentation masks), the automatic cell classification system, or by a manual inspection (see evaluation of segmentation results).

Splitting clusters and constructing object masks. The markers in the marker frame image were used to extract subimages of individual nuclei or clusters of nuclei from the original shade corrected frame image (which was not median-filtered). All the markers in the marker image were labeled. This was done by computing a 2D label matrix L_m where the lth marker included all elements in L_m that had value l. Markers which included frame boundary pixels were excluded. Based on the bounding box of each marker (with size x_{size}, y_{size}), a subimage corresponding to a larger rectangle was extracted from the frame image, see Figure 3a. The size of the subimage [x_{size} + $2(r_s + 20)$, $y_{size} + 2(r_s + 20)$] was chosen such that it was large enough to contain the reproduced original object mask (Fig. 3b) and also a larger mask that was used for finding a start contour for the snake (see below and Fig. 3f).

If a part of the larger rectangle was outside the frame image the size of the rectangle was reduced. A corresponding subimage was extracted from the marker image. Using information from the marker label matrix L_m, markers corresponding to other nuclei (i.e., markers corresponding to elements in $L_m \neq 1$) were removed from the marker subimage. The convex hull solidity of each marker was computed and used to identify markers corresponding to single nuclei (solidity $\geq$0.9) and markers corresponding to two or more nuclei (solidity $<$
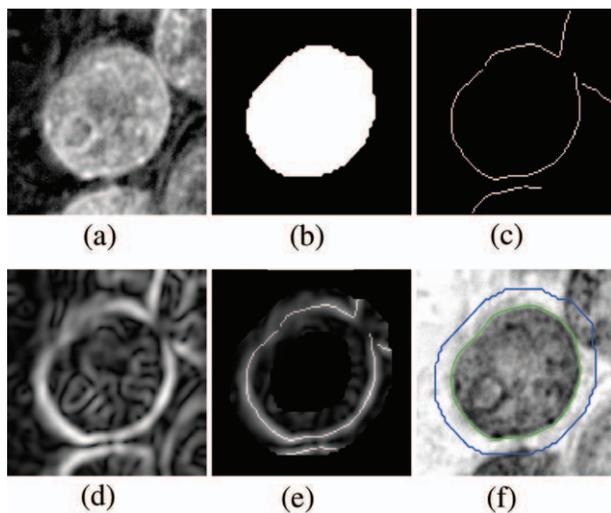
Figure 3. (**a**) A gray scale subimage extracted from the frame image. (**b**) The object mask reproduced from a marker by dilation. (**c**) The Canny binary edge map. (**d**) The gradient magnitude. (**e**) A bounded, annular edge map. (**f**) The initial contour for the snake (blue) and the final snake segmentation result (green).

0.9). The solidity value of 0.9 was chosen by trial and error on our training frames.

To split a marker corresponding to more than one nucleus, the marker was eroded iteratively with a 3 × 3 plus-shaped structuring element. To avoid deleting the marker, the iteration was performed until the marker was split into two or more markers, or until the marker area was ≤15 pixels. Each of the split markers that had a solidity ≥0.6 was regarded as a marker corresponding to a single nucleus, and a new marker subimage was created where markers corresponding to other nuclei were removed. The split markers generally had much lower solidity values than 0.9, and a solidity criterion of 0.6 was chosen by trial and error to avoid missing candidate nuclei. Split markers with a solidity value <0.6 were rejected.

To reproduce an object mask from a marker, the marker was dilated with the same structuring elements as were used for producing the marker (Fig. 3b). If >25% of the reproduced object mask overlapped with a mask previously stored in the image "reproduced object masks" (as described below), then the nucleus was regarded as already segmented in a previous iteration.

Optimizing the initial contours for the nuclear snakes. If the nucleus corresponding to the present marker was not segmented yet, then a larger and a smaller mask were produced, to prepare for the final segmentation using a snake. To produce a larger mask, the reproduced object mask was dilated with a disk shaped structuring element with a radius of 11 pixels. The boundary of the larger mask was used as a start contour for the snake (see Fig. 3f). To produce a smaller mask, the reproduced object mask was eroded iteratively with a 3 × 3 plus-shaped structuring element until either the maximum number of iterations (= 15) was reached or as long as the area of the smaller mask was > 15 pixels.

Optimizing the convergence of the nuclear snakes. The Canny edge detector (36,38) was applied on the gray scale subimage. The gradient magnitude was computed using the derivative of a Gaussian. The standard deviation of the Gaussian filter was set to $\sigma = 2.0$ and the low and high gradient threshold values were set to 0.18 and 0.7, respectively. The resulting Canny binary edge map (Fig. 3c) was added to the gradient magnitude (Fig. 3d). By adding the gradient magnitude to the binary edge map, we may obtain some gradient information in areas where the binary edge map was missing edge information. Some gradient information was then removed from this image based on the larger and the smaller mask. All pixel values corresponding to pixels with value 0 in the "larger mask" binary image was set to 0 and all pixel values corresponding to pixels with value 1 in the "smaller mask" binary image was also set to 0, resulting in a bounded, annular edge map (Fig. 3e).

Final segmentation of each nucleus using the GVF snake. The GVF snake was used to perform the final segmentation of each nucleus (Fig. 3f). The GVF of the bounded annular edge map was computed. The GVF regularization coefficient μ was set to 0.2 and the number of GVF iterations was set to 80. The GVF was then normalized and used as an external force field. The points in the snake start contour were interpolated to have equal distances. The desired resolution between the points was set to 0.8. The internal force parameters were set as follows: elasticity $\alpha = 5$, rigidity $\beta = 10$, viscosity $\gamma = 1$. The external force weight was set to $\kappa = 0.6$. The snake deformation was iterated 30 times. After each deformation the snake was interpolated to have equal distances (with the desired resolution between the points set to 0.8).

Saving the segmentation masks. A mask was created for each segmented nucleus using the Matlab function poly2mask (36). A nucleus was considered (by trial and error on the training frames) to be successfully segmented if the segmentation mask satisfied the following criteria:

- The solidity of the mask was greater or equal to 0.97
- The average gradient value along the boundary of the mask was >0.3
- The area of the mask was >450 and <16,000 pixels

In the present study, the solidity of the segmentation masks of successfully segmented nuclei was very high. In other materials with more misshapen cancer nuclei (e.g., colon and gynecological cancers), the solidity criterium should be set to lower values. If the nucleus was successfully segmented then (i) the reproduced object mask was stored in a binary image "reproduced object masks" and (ii) the final segmentation mask was stored in a "segmentation mask image" as a labeled connected component. If there was an overlap between the current segmentation mask and a previous stored mask, then the pixel values of the pixels in the overlapping area were set to an arbitrary high value to indicate that this area belongs to both segmentation masks.
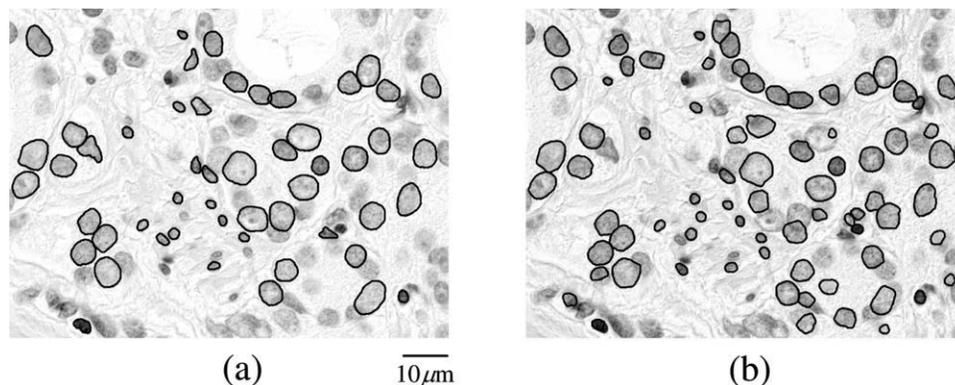
Figure 4. Result from (**a**) manual and (**b**) automatic segmentation of a training frame image from ''Case 1.''

Splitting the overlapping masks. An overlap between the present segmentation mask and a previous defined segmentation mask of <5% of the area of the present mask was accepted. If the overlap was between 5 and 25%, the overlapping masks were split. Two overlapping segmentation masks were split by first identifying the two points where the boundaries of the two masks intersected and then drawing a straight line between these two points. The Bresenham line algorithm (39) was used to determine which pixels in the 2D image that should be included to form a close approximation to a straight line between two given points.

Evaluation of Segmentation Results

The development of an automatic cell nucleus segmentation procedure to replace manual segmentation is motivated by the need to save time and cost in handling increasing volumes of images. But an overriding concern is the agreement between the two methods. Can we use segmentation results from the manual and automatic approaches interchangeably?

There are several aspects to consider:

- Does the automatic algorithm segment the same (number of) cell nuclei as the experts do?
- For the manually segmented nuclei that are also segmented by the automatic method, we need to inspect the accuracy of the segmentation result.
- For the nuclei that are only segmented by one of the methods, we need to check if there are any systematic differences in typical nuclear parameters.

We use a single manual delineation of selected nuclei as a surrogate of truth (31). Trained personnel (two experts) performed the manual segmentation using a drawing table and special-written software. Starting with the first frame from each case, all in-focus, non-overlapping complete nuclei were selected from consecutive frames (see Fig. 4a and Supporting Information). As we are measuring the amount of light absorbed by the nuclear staining, overlapping nuclei must be avoided to ensure correct measurements per nucleus. The instruction was "draw a line just outside the nuclei", i.e., the area lying inside the

contour was regarded as the nucleus. The precision may depend on several factors, e.g., personal experience, speed of drawing and software used. We observed no systematic difference in the manual segmentations performed by the two experts. If >5% of the area of a manual segmentation mask overlapped with another manual segmentation mask, the overlapping nuclei were regarded as incorrect manual segmentations.

Image-based and object-based evaluation. In supervised evaluation, the difference between a reference segmentation and the output of a segmentation algorithm is quantified. However, there are two different classes of such evaluations: image-based and object-based, quantifying the performance of the method per image or per object, respectively. This results in two different uses of terms like "under-" and "over-segmentation", as well as "sensitivity" and "specificity."

In image-based evaluation the concept of "under-segmentation" means that one automatically segmented region covers several nuclei from the ground-truth (also called "degeneracy"), while "over-segmentation" means that some nuclei have been split into more than one object, and the concept of "sensitivity" is often taken to mean "the percentage of hand-segmented objects that were automatically segmented with good quality." Thus, the geometrical accuracy of the segmentation of each object is often not considered.

In object-based evaluation we often differentiate between three classes of metrics: probabilistic or pixel-based, edge-based, and feature based. Probabilistic evaluation is based on all the pixels within the "true" (manual) segmented object and the segmentation obtained by an automatic method, respectively. The concept of "under-segmentation" means the fraction of the manual segment that is missed by the automatic method, while "over-segmentation" is the set of pixels falsely segmented as object, as a fraction of the true object area. We may also use the statistical concepts of sensitivity and specificity to characterize the segmentation performance per object, where "sensitivity" is the fraction of the true object that is actually segmented as object pixels, while "specificity" is the fraction of the true background that is segmented as background pixels. This conforms with the use of these terms in Refs. 30, 34, and 35.

In the present article we will not consider image-based evaluation, both because we do not have problems with degeneracy (handled by the solidity measure) and splitting, and because we need to focus on the detailed accuracy of the segmentation of each object, as our rationale for automatic segmentation is the further detailed analysis of the nuclear interior.

Assessment of accuracy for each nucleus. Accuracy denotes the degree to which the segmentation result agrees with the truth (31). For each nucleus that was both manually and automatically segmented, the true (manual) delineation was represented by a binary image $M(x,y)$, where the object (nucleus) was represented by pixels with value $M(x,y) = 1$, while the background was represented by pixels with value $M(x,y) = 0$. This subimage was compared with a corresponding binary subimage, $A(x,y)$, of equal size containing the mask obtained by automatic segmentation where the object and background pixels had values $A(x,y) = 1$, and $A(x,y) = 0$, respectively.

Each pixel in the image of a given nucleus and its surroundings was classified according to their binary M- and A- value, and the number of true positive (TP), false negative (FN), false positive (FP), and true negative (TN) pixels were counted:

$$\begin{array}{llll} \text{TP} & \text{if} & M(x,y) = 1 \wedge A(x,y) = 1 \\ \text{FN} & \text{if} & M(x,y) = 1 \wedge A(x,y) = 0 \\ \text{FP} & \text{if} & M(x,y) = 0 \wedge A(x,y) = 1 \\ \text{TN} & \text{if} & M(x,y) = 0 \wedge A(x,y) = 0 \end{array} \qquad (4)$$

where $\wedge$ is the logical *and*-operator.

The undersegmentation is given by US = FN/(TP + FN). The oversegmentation is given by OS = FP/(TP + FN). The sensitivity of the automatic segmentation method is the portion of the pixels segmented as nuclear pixels of all the nuclear pixels in the manual segmentation: $p = TP/(TP+FN)$.

The specificity of the automatic segmentation method is the portion of the pixels segmented as background of all the background pixels in the manual segmentation: $q = TN/(TN+FP)$. A receiver operating characteristic (ROC) is a graphical plot of the sensitivity versus (1-specificity) (40).

The measures above do not include information about the spatial distribution of the errors, the connectedness of the over- and under-segmentation regions, or the distance from the perimeter of the true segment to a given erroneously segmented pixel.

If the under-segmentation was >0.25, the automatic segmentation was regarded as an incorrect segmentation (e.g., a segmentation of an object within the nucleus) and the nucleus was regarded as not detected by the automatic segmentation

Number of segmented nuclei and nuclear features. For each frame the number of manually and automatically segmented nuclei was counted. The percentage of manually segmented nuclei that were also segmented by the automatic method was computed. This was done automatically by comparing the binary mask images from the manual and automatic segmentation.

For both manually and automatically segmented nuclei, the nuclear area was recorded. Because even small segmentation errors may influence both the measured mean gray level and the standard deviation of the nuclear gray level, as well as estimated textural parameters within the nuclear area, we also recorded the mean and standard deviation of the pixel gray level values, and the first order entropy (computed from the nuclear gray level histogram) for each nucleus.

Evaluation of the method on four cases of prostate cancer in another study. After automatic nuclear segmentation, the next step in our nucleotyping system is an automatic cell nucleus classification system. This classifies each segmented object into one of the following categories/galleries; (1) epithelial cancer cell nuclei, (2) blood cell nuclei, (3) fibroblast cell nuclei, or (4) noncell nuclei. After automatic cell nuclei classification, the galleries are manually inspected and edited by human experts and the resulting Gallery 1 of each case includes all epithelial cancer cell nuclei that are of good enough quality for nuclear texture analysis.

As an extraquality check of the robustness of our segmentation method, we have manually inspected the segmentation of nuclei that were included in Gallery 1 when the C++ version of our segmentation software and the automatic cell classification software was applied on 260 cases of prostate cancer in another study. We selected four random cases and some random frames from each of these cases, and counted the number of nuclei that were segmented with "good quality" for further nuclear texture analysis. The outcome of this experiment is reported in the Results section.

RESULTS

Two frame images were used to develop the segmentation method, and the first frame is used as an "example frame" in some of the figures (see Fig. 4). The method was tested on 28 randomly selected frame images from three cases (11 frames from Case 1, 8 from Case 2, and 9 from Case 3).

Manual delineation of the 30 frames resulted in a total of 924 manually segmented cell nuclei, of which 4 were overlapping nuclei and therefore excluded from the evaluation below. From the first part of Table 1 we see that if we include all nuclei larger than 450 pixels, about 68% of the manually segmented nuclei were also segmented by the automatic method. The mean sensitivity (95%) and the mean specificity (96%) averaged over all the nuclei segmented both manually and automatically in the 30 frames from the three cases is (almost) the same as in the training frames. Figure 5 shows ROC data for each frame of the three cases. The automatic method segmented 96% more "nuclei" compared to the manual method. However, some of the "nuclei" that were segmented only by the automatic method were erroneous segmentations capturing only a part of the nucleus.

If we include only the biologically most relevant nuclei, i.e., the nuclei larger than 2,400 pixels (corresponding to a nuclear diameter $D > 5.6$ μm), the automatic method segmented only 13% more candidates for nuclei compared to the manual method, about 73% of the manually segmented nuclei were

Table 1. The percentage of manually segmented nuclei that were also segmented by the automatic method, the mean sensitivity and specificity of the automatically produced segmentation masks, and the total number of automatically and manually segmented nuclei

	TRAINING	CASE 1	CASE 2	CASE 3
Nuclear area > 450 pixels				
% of manual segmentations	73.4	71.3	68.7	64.9
Sensitivity	95.5	94.8	94.6	94.8
Specificity	96.5	96.0	96.4	95.5
Automatic/manual (ratio)	144/96 (1.5)	770/426 (1.81)	512/230 (2.23)	530/268 (1.98)
Nuclear area > 2400 pixels (diameter > 5.6 μm)				
% of manual segmentations	83.0	75.1	75.2	68.1
Sensitivity	95.3	95.0	95.6	95.5
Specificity	96.0	96.3	96.5	95.5
Automatic/manual (ratio)	73/61 (1.2)	373/328 (1.14)	186/172 (1.08)	239/206 (1.16)

The results are given for nuclei having an area > 450 pixels (top) and for nuclei having an area > 2,400 pixels (bottom).

also segmented by the automatic method, but the mean segmentation sensitivity/specificity remained at 95%/96% (second part of Table 1).

Comparing Nuclear Area of Nuclei Segmented by Both Methods

A correlation analysis indicates a very close relation between the nuclear area of the manual and automatic segmentation results on the same nuclei ($r = 0.99$). However, the high correlation may be quite misleading, and Bland–Altman scatter plots (41) of feature differences versus feature mean values are better suited to compare the results from nuclei segmented by both approaches.

In the Bland–Altman plot (Fig. 6), the estimated mean difference μ_d reveals that there is a small but systematic bias towards a slightly larger manual nuclear area. From Figure 6 we can also observe that the spread of the difference varies with the area, so that the "limits of agreement" (LOA $= \mu_d \pm 2\sigma_d$) of the area difference, and the 95% confidence intervals of these limits, cannot be used to assess whether the area accuracy is acceptable or not. However, the curves in Figure 6 illustrate that a large fraction (60–80%) of the spread seen in the Bland–Altman plot is within the error margin of a one pixel wide perimeter outside/inside each manually segmented object. So the area accuracy seems to be acceptable for both small and large nuclei.

Comparing Nuclear Gray Level Features of Nuclei Segmented by Both Methods

An interesting question is whether the observed bias and spread in nuclear area is reflected in various important nuclear features (see Fig. 7). The regression line in Figure 7a is given by DM(g) $= 0.004 + 0.097$ (±0.021) DM(a), with a correlation coefficient of $r = 0.47$. This implies that a 6% difference in area compared to the mean (which covers about 75% of the spread in the Bland–Altman area plot) will result in a mere 1% change in the difference-to-mean ratio of the mean nuclear gray level.

The regression line in Figure 7b is given by DM(H) $= 0.008 + 0.119$ (±0.017) DM(a), with $r = 0.62$. This implies

that a 6% difference in area compared to the mean will result in only a 1.5% change in the difference-to-mean ratio of the nuclear entropy value. If the automatic segmentation is close to the real nuclear boundary, while the manual one is slightly larger, the latter will include some pixels from the brighter background, increasing the nuclear entropy. The difference-to-mean ratio of the gray level power-to-mean ratio (PMR = standard deviation divided by mean value), which is highly correlated with entropy, gave similar results.

Comparing Nuclear Features of Nuclei Segmented by Only One of the Methods

The biologically most relevant nuclei, i.e., the nuclei larger than 2,400 pixels, were split into three groups (Supporting Information):

- Group 1: Cell nuclei segmented by both methods (already analyzed)
- Group 2: Cell nuclei segmented only by manual experts
- Group 3: Cell nuclei segmented only by the automatic algorithm

We have checked if there are any systematic differences in characteristic nuclear parameters (e.g., nuclear area, mean nuclear gray level, and gray level entropy) between Group 1 and 2, and between Group 1 and 3. The differences between parameters obtained from the manual and automatic segmentation of nuclei of Group 1 have been analyzed above for all sizes of nuclei.

We have investigated whether the distributions in the two scattergrams: (a) mean nuclear gray level versus nuclear area, (b) gray level entropy versus nuclear area, are similar for the manual and the automatic segmentation of the group of nuclei that have been segmented both manually and automatically (Group 1), and whether there are differences between the scattergrams from manual segmentation in Group 2 and 1, and from Group 3 and 1.

We found that for Group 1 (the nuclei segmented by both methods), the difference between the manual and the automatic segmentation results are within acceptable limits. The comparison with Group 2 (the nuclei segmented only
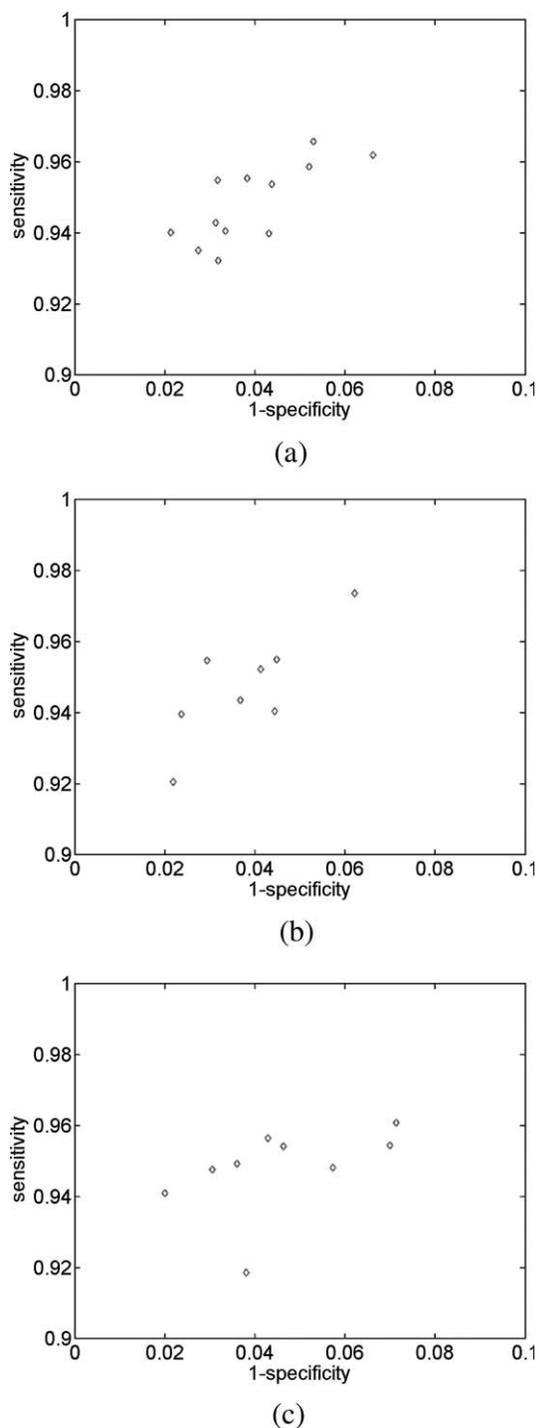
(a)

(b)

(c)

Figure 5. ROC data computed from 13 image frames from ''Case 1'' (**a**), eight frames from ''Case 2'' (**b**), and nine frames from ''Case 3'' (**c**). The segmentation sensitivity on the vertical axis, and (1-specificity) on the horizontal axis are the mean values computed from all nuclei of the frame that were both manually and automatically segmented. The graphs show only the 1% upper left corner of the ROC plane.

manually) shows that the distribution of nuclei that are left out by the automatic segmentation algorithm is not biased with respect to area, average gray level or gray level entropy.

Group 3 (the nuclei segmented only by the automatic algorithm) on the other hand, seems to be biased towards smaller area and a slightly larger range of mean nuclear gray level and gray level entropy.

An Expert Visual Evaluation

As a separate test, we have performed an expert visual inspection of both the manual and the automatic segmentations of the biologically most relevant nuclei (area > 2,400 pixels), where the 25% threshold on under-segmentation was not in effect. This corresponds to a situation where the "ground truth" is not available for comparison with the result from the automatic segmentation algorithm. About 2.5% of the manual segmented nuclei were judged to be erroneous, whereas 12% of the automatic segmented nuclei were erroneous, resulting in a few more correctly segmented nuclei compared to the manual method.

Evaluation of the Method on Four Cases of Prostate Cancer in Another Study

The manual inspection of the segmented nuclei of four cases of prostate cancer in another study showed that between 93 and 98% of the epithelial cancer cell nuclei were segmented with good quality (Table 2). We regard 2–7% erroneous segmented nuclei as a very good result. Depending on the downstream use in the nucleotyping system one may choose to manually exclude these nuclei before further analysis, or go with the old principle "do more less well" allowing for an almost unlimited number of nuclei as manual editing is not required.

The nucleotyping system typically measured about 2,500–3,000 frames/case and a maximum limit of 5,000 frames was set for the largest cases, resulting in about 30,000 nuclei/case. The segmentation of a typical frame image was performed in 2 min, and using our cluster consisting of 192 kernels about 200 frames were segmented in 2 min, and all the frames of the 260 cases were segmented in about 5 days.

Robustness of the Method to Changes in Parameter Values

The result of the proposed segmentation method depends on a sequence of image processing steps containing a large number of parameter values. We have analyzed the importance of including each step and the robustness or stability of the method by changing one parameter value at a time, keeping the other parameter values constant. These tests were performed on the image frames from Case 1, considering all nuclei >450 pixels.

The adaptive thresholding of the gray level frame images is clearly important (Fig. 2a) and the postprocessing step of Yanowitz and Bruckstein (27) is crucial in removing false objects from the binary frame image (Fig. 2b). A δ value of 0.2 (in the Niblack method) gave good results, but changing this to 0.1 or 0.3 did not influence the sensitivity or specificity. The object validation step of Yanowitz and Bruckstein involves a threshold on the gradient along the perimeter of the object
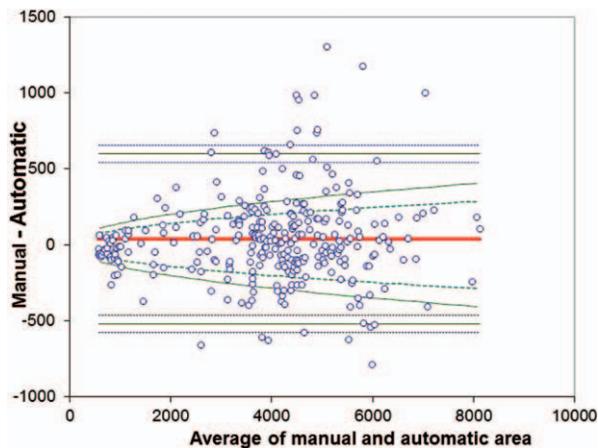
Figure 6. Bland-Altman plot of the difference in nuclear area between manually and automatically segmented nuclei from "Case 1" versus the average area from the two methods. The mean difference μ_d (red solid line), the "limits of agreement" LOA (green solid lines), and the 95% confidence intervals of the LOA (blue dashed lines) are also shown. The curves show the difference-versus-average of digitized circular objects with or without an outside/inside 4- (dashed curves) or 8-connected perimeter (solid curves). The smallest nuclei are regarded as blood cell nuclei, while the larger nuclei (with more than 2,400 pixels or a nuclear diameter $D > 5.6\ \mu m$) are regarded as epithelial cancer cell nuclei.

candidate (here chosen as 0.1). A 50% change in this parameter did not influence the sensitivity or specificity.

Another crucial component of the method is the GVF snake. Initial tests using a traditional snake (21) required a very precise initialization, and therefore gave poor results. Using the GVF snake with an annular edge map, on the other hand, produced very satisfactory results. Halving the elasticity and rigidity parameters ($\alpha = 5$, $\beta = 10$) reduced the number of detected nuclei from 71 to 67% of the manually segmented nuclei, but did not affect sensitivity and specificity. Reducing the external force weight from $\kappa = 0.6$ to 0.5 increased the number of detected nuclei from 71 to 73% of the manually segmented nuclei, but again this did not affect sensitivity or specificity. The number of iterations (of the snake deformation) was set to 30. Increasing the number of iterations did not improve the results, but the same results were obtained by just 10 iterations. The GVF regularization coefficient, on the other hand, will influence the results. It was set to $\mu = 0.2$, and increasing it to $\mu = 0.3$ brought the sensitivity down from 95 to 94%, the specificity down from 96 to 89%, and the number of detected nuclei from 71 to 60% of the manually segmented nuclei. Decreasing it to $\mu = 0.1$ lowered the number of detected nuclei from 71 to 69% of the manually segmented nuclei but did not affect the sensitivity or specificity.

DISCUSSION

We have developed a method for automatic segmentation of cell nuclei from Feulgen stained tissue sections of prostate cancer. The algorithm was developed on a few frames and tested on a total of 30 frames from three patients. Comparing
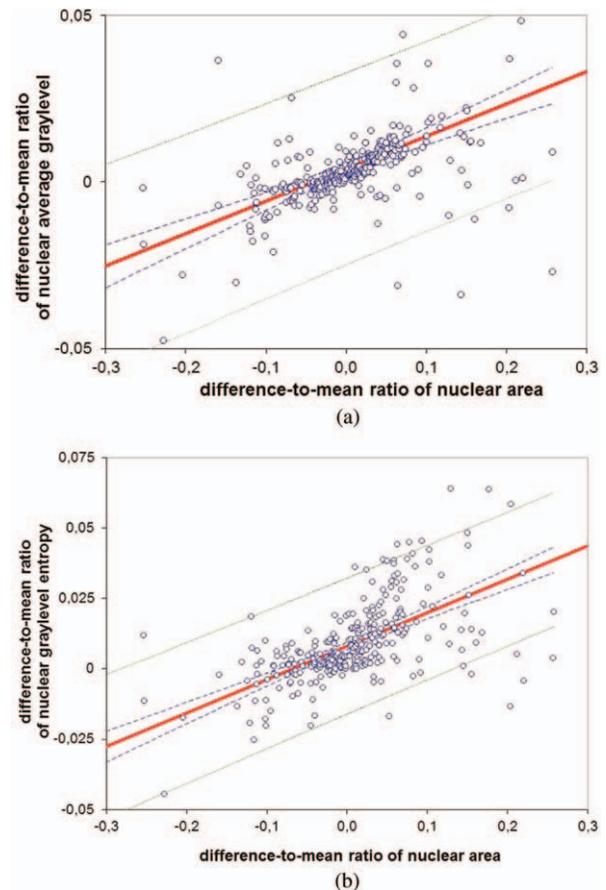

Figure 7. The difference-to-mean ratio of two nuclear features [(**a**) the mean gray level and (**b**) the gray level entropy)] for the manual and automatic segmentation versus the corresponding difference-to-mean ratio of the nuclear area for the nuclei of "Case 1." The regression line (red), the 95% confidence interval of the regression line (blue dashed), and the 95% prediction interval (green dashed) are also given.

the results from the automatic algorithm to the manual segmentation of the same set of frames, the automatic method segmented a few more nuclei compared to the manual

Table 2. The automatic segmentation was applied on 260 cases of prostate cancer in another study

CASE	NO. FRAMES	NO. NUCLEI	% VERY ACCURATE	% ACCEPTABLE	% ERROR
Case 1	45	664	88.9	93.1	6.9
Case 2	49	1415	90.6	93.0	7.0
Case 3	56	920	91.0	98.0	2.0
Case 4	51	953	86.6	95.4	4.6

The segmentation of nuclei that were classified as epithelial cancer cell nuclei by the automatic cell classification system was manually inspected. About 3,952 nuclei from 201 random frames from 4 random cases were judged as; (i) very accurate segmentation, (ii) segmented with a small, but acceptable error, (iii) segmented with a large, unacceptable error.

method, but about 73% of the manually segmented nuclei were also segmented by the automatic method.

A total of 924 cell nuclei were both manually and automatically segmented. An accuracy analysis indicated a very close correlation between the area of the manual segments and the areas of the automatic segmentation results on the same nuclei ($r = 0.99$). The mean sensitivity was 95% and the mean specificity was 96%. There was, however, a systematic bias towards a slightly larger manual nuclear area, and a spread consistent with a 1–2 pixel wide perimeter outside/inside the manually segmented nuclei. Scattergrams of characteristic nuclear parameters have verified that for the nuclei segmented by both methods, the difference between the manual and the automatic segmentation results are within acceptable limits.

We have also verified that the distribution of nuclei that are left out by the automatic segmentation algorithm (i.e., the nuclei segmented only by manual experts) is not biased with respect to area, mean gray level or gray level entropy. The distribution of nuclei segmented only by the automatic algorithm, on the other hand, seems to be biased towards smaller area and a somewhat larger range of mean nuclear gray level and gray level entropy.

The objects that were segmented only by the automatic method could be: (i) correctly segmented nuclei, (ii) incorrectly segmented nuclei, or (iii) false objects ("nonnuclei"). Segmented objects that should not be included in further texture analysis are handled by the automatic cell classification system and the manual inspection performed after automatic segmentation and cell classification. An independent test of the method on four cases in another study showed that between 93 and 98% of the epithelial cancer cell nuclei were segmented with good enough quality for further texture analysis.

The reproducibility of both the manual and the automatic segmentations remains an interesting question. This could be studied by letting more individuals do manual segmentations of the same image fields, and repeating automated segmentation of the same microscopic fields digitized multiple times, assessing eventual statistical differences between multiple manual and computer segmentations. Such a task is however, a whole new study, and we have therefore not attempted to pursue this line in the present article.

A quantitative evaluation of the segmentation results is vitally important. This is true not only because we need a quantified agreement to replace manual segmentation with an automatic procedure. It is definitely needed if we are to prove the validity of a given method based on a data set that is much too large to oversee and evaluate manually. Thus, the assessment of segmentation accuracy as well as the intercomparison of the three groups of object parameter distributions should be streamlined and incorporated into the segmentation algorithm. Then, the testing and resetting of algorithm parameters for a convergence towards an optimal segmentation result could be more or less automatic. Ideally, the evaluation of accuracy and bias could be constructed as a feedback mechanism, using an automatic gradient search for an optimal set of parameters to the segmentation algorithm. But even at the present stage, we believe that our approach opens the possibil-

ity of large-scale nuclear feature analysis based on an automatic segmentation of cell nuclei in Feulgen stained histological sections.

Literature Cited

1. Nielsen B, Albregtsen F, Danielsen HE. Statistical nuclear texture analysis in cancer research: A review of methods and applications. Crit Rev Oncogenesis 2008;14:89–164.
2. Bengtsson E, Wahlby C, Lindblad J. Robust cell image segmentation methods. Pattern Recogn Image Analysis 2004;14:157–167.
3. Bedin V, Adam RL, de Sa BCS, Landman G, Metze K. Fractal dimension of chromatin is an independent prognostic factor for survival in melanoma. BMC Cancer 2010;10:260.
4. Montironi R, Scarpelli M, Lopez-Beltran A, Mazzucchelli R, Alberts D, Ranger-More J, Bartels HG, Hamilton PW, Einspahr J, Bartels PH. Chromatin phenotype karyometry can predict recurrence in papillary urothelial neoplasms of low malignant potential. Cell Oncol 2007;29:47–58.
5. Axelrod DE, Miller NA, Lickley HL, Qian J, Christens-Barry WA, Yuan Y, Fu Y, Chapman J-AW. Effect of quantitative nuclear image features on recurrence of ductal carcinoma in situ (DCIS) of the breast. Cancer Inform 2008;6:99–109.
6. Korde VR, Bartels H, Barton J, Ranger-More J. Automatic segmentation of cell nuclei in bladder and skin tissue for karyometric analysis. Anal Quant Cytol Histol 2009;31:83–89.
7. Wahlby C, Sintorn I-M, Erlandsson F, Borgefors G, Bengtsson E. Combining intensity, edge and shape information for 2D and 3D segmentation of cell nuclei in tissue sections. J Microsc 2004;215:67–76.
8. Al-Kofahi Y, Lassoued W, Lee W, Roysam B. Improved automatic detection and segmentation of cell nuclei in histopathology images. IEEE Trans Biomed Eng 2010;57:841–852.
9. Quelhas P, Marcuzzo M, Mendonca AM, Campilho A. Cell nuclei and cytoplasm joint segmentation using the sliding band filter. IEEE Trans Med Imaging 2010;29:1463–1473.
10. Adiga U, Malladi R, Fernandez-Gonzalez R, Ortiz de Solorzano C. High-throughput analysis of multispectral images of breast cancer tissue. IEEE Trans Image Process 2006;15:2259–2268.
11. Dufour A, Shinin V, Tajbakhsh S, Guillen-Aghion N, Olivo-Marin JC, Zimmer C. Segmenting and tracking fluorescent cells in dynamic 3-D microscopy with coupled active surfaces. IEEE Trans Image Process 2005;14:1396–1410.
12. Li G, Liu Y, Tarokh A, Nie J, Guo L, Mara A, Holley S, Wong STC. 3D cell nuclei segmentation based on gradient flow tracking. BMC Cell Biol 2007;8:40.
13. Li F, Zhou X, Ma J, Wong STC. Multiple nuclei tracking using integer programming for quantitative cancer cell cycle analysis. IEEE Trans Med Imaging 2010;29:96–105.
14. Gudla PR, Nandy K, Collins J, Meaburn KJ, Misteli T, Lockett SJ. A high-throughput system for segmenting nuclei using multiscale techniques. Cytom A 2008;73A:451–466.
15. Chen C, Li H, Zhou X, Wong STC. Constraint factor graph cut-based active contour method for automated cellular image segmentation in RNAi screening. J Microsc 2008;230:177–191.
16. Cheng J, Rajapakse JC. Segmentation of clustered nuclei with shape markers and marking function. IEEE Trans Biomed Eng 2009;56:741–748.
17. Hodneland E, Bukoreshtliev NV, Eichler TW, Tai XC, Gurke S, Lundervold A, Gerdes HH. A unified framework for automated 3D segmentation of surface-stained living cells and a comprehensive segmentation evaluation. IEEE Trans Med Imaging 2009;28:720–738.
18. Baggett D, Nakaya MA, McAuliffe M, Yamaguchi TP, Lockett S. Whole cell segmentation in solid tissue sections. Cytom A 2005;67A:137–143.
19. McCullough DP, Gudla PR, Harris BS, Collins JA, Meaburn KJ, Nakaya MA, Yamaguchi TP, Misteli T, Lockett SJ. Segmentation of whole cells and cell nuclei from 3D optical microscope images using dynamic programming. IEEE Trans Med Imaging 2008;27:723–734.
20. Malpica N, de Solórzano CO, Vaquero JJ, Santos A, Vallcorba I, Garcia-Sagredo JM, del Pozo F. Applying watershed algorithms to the segmentation of clustered nuclei. Cytometry 1997;28:289–297.
21. Kass M, Witkin A, Terzopoulos D. Snakes—Active contour models. Int J Comput Vis 1988;1:321–331.
22. McInerney T, Terzopoulos D. Deformable models in medical image analysis: A survey. Med Image Anal 1996;1:91–108.

23. Plissiti ME, Nikou C, Charchanti A. Combining shape, texture and intensity features for cell nuclei extraction in Pap smear images. Pattern Recog Lett 2011;32:838–853.

24. Trier ØD, Taxt T. Evaluation of binarization methods for document images. IEEE Trans Pattern Anal Machine Intell 1995;17:312–315.

25. Trier ØD, Jain AK. Goal-directed evaluation of binarization methods. IEEE Trans Pattern Anal Machine Intell 1995;17:1191–1201.

26. Niblack W. An Introduction to Digital Image Processing. Birkeroed, Denmark: Strandberg Publishing Company; 1986.

27. Yanowitz SD, Bruckstein AM. A new method for image segmentation. Computer Vis Graph Image Process 1989;46:82–95.

28. Xu C, Prince JL. Snakes, shapes, and gradient vector flow. IEEE Trans Imag Proc 1998;7:359–369.

29. Xu C, Prince JL. Gradient vector flow (GVF) active contour toolbox. Available at: http://iacl.ece.jhu.edu/projects/gvf

30. Udupa JK, LaBlanc VR, Schmidt H, Imielinska C, Saha PK, Grevera GJ, Zhuge Y, Molholt P, Jin Y, Currie LM. Methodology for evaluating image segmentation algorithms. Proc SPIE Med Imaging 2002;4684:266–277.

31. Udupa JK, LaBlanc VR, Zhuge Y, Imielinska C, Schmidt H, Currie LM, Hirsch BE, Woodburn J. A framework for evaluating segmentation algorithms. Comput Med Imaging Graph 2006;30:75–87.

32. Warfield SK, Zou KH, Wells WM. Validation of image segmentation and expert quality with an expectation–maximizing algorithm. In: Dohi T, Kokinis R, editors. Proceeding of MICCAI 2002, the Fifth International Conference. Springer, Heidelberg, pp 298–306.

33. Warfield SK, Zou KH, Wells WM. Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation. IEEE Trans Medical Imaging 2004;23:903–921.

34. Popovic A, de la Fuente M, Engelhardt M, Radermacher K. Statistical validation metric for accuracy assessment in medical image segmentation. Int J Comput Assisted Radiol Surg 2007;2:169–181.

35. Einstein AJ, Gil J, Wallenstein S, Bodian CA, Sanchez M, Burstein DE, Wu H-S, Liu Z. Reproducibility and accuracy of interactive segmentation procedures for image analysis in cytology. J Microsc 1997;188:136–148.

36. Image Processing ToolBox User's Guide, Version 5, The MathWorks, Inc. 2004.

37. Soille P. Morphological Image Analysis: Principles and Applications. Springer-Verlag; 1999. pp 173–174.

38. Canny J. A computational approach to edge detection. IEEE Trans Pattern Analysis Machine Intell 1986;8:679–714.

39. Bresenham JE. Algorithm for computer control of a digital plotter. IBM Systems J 1965;4:25–30.

40. Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recogn 1997;30:1145–1159.

41. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307–310.