Quantification of epithelial area by image processing applied to endometrial carcinoma: a comparison with ovarian tumors
Schipper, N.W.; Smeulders, A.W.M.; de Lange, J.H.M.; Baak, J.P.A.

Published in:
Human Pathology

DOI:
10.1016/0046-8177(89)90233-5

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Quantification of Epithelial Area by Image Processing Applied to Endometrial Carcinomas: A Comparison With Ovarian Tumors

NELLIE W. SCHIPPER, MSc, ARNOLD W. M. SMEULDERS, MSc, PhD,* JOS H. M. DE LANGE, AND JAN P. A. BAAK MD, PhD

In endometrial carcinomas, the epithelial area measured by interactive morphometry is an important feature in the classification of tumors of varying histologic grades. This report describes an image analysis technique for the fully automated estimation of the area percentages of epithelium and stroma in tissue sections of the endometrium obtained from hysterectomy specimens. The method is evaluated using endometrial carcinomas with varying degrees of malignancy. From standard paraffin sections stained with pararosanilin Feulgen and naphthol yellow, a blue-yellow image pair was recorded. The blue image was used to determine the total tissue area and the yellow image was used to determine the epithelial area. Image processing of the blue image was comprised of correction for shading, segmentation of the tissue area, and restoration of the segmented image by removing small artefacts and closing small tears in the tissue. Image processing of the yellow image was based on the fact that epithelial nuclei are generally more tightly packed than stromal nuclei and consists of the following steps: correction for shading, gaussian blurring, segmentation of nuclei, and editing the segmented image by removing small objects and closing small spaces between the epithelial nuclei. These image processing steps are compared with those used for quantification of the epithelial area in ovarian tumors. The performance of the method was evaluated using 120 image pairs from 30 endometrial carcinomas of varying histologic grades. The epithelial area percentages, as assessed by digital image processing, strongly correlate to control percentages that were established by interactive morphometry (r = .987). HUM PATHOL 20:1125-1132. © 1989 by W.B. Saunders Company.

Despite the fact that the overall 5-year survival rates for patients with endometrial carcinoma are relatively good compared with those for patients with other gynecologic tumors,1,2 20% to 30% of all stage I patients die from distant metastases. Therefore, objective predictors of clinical outcome are needed. Histologic type is associated with the prognosis, but most cases of endometrial carcinoma are of the common adenocarcinoma type. Nuclear and histologic grade have prognostic value; however, assigning these grades is not always easily reproducible.3,4 There may be considerable variation among the assessments of several pathologists which may result in undertreatment or overtreatment of an individual patient. Therefore, objective and more reproducible features are required to prevent these undesired consequences of falsely positive and falsely negative diagnoses.

In one study,7 the epithelial area percentage has been found to be an important discriminator between endometrial carcinomas of different grades and may thus have therapeutic significance. In many institutions, moderately and poorly differentiated cancer patients (grades II and III) receive postoperative radiotherapy, in contrast to the treatment given to patients with a well-differentiated carcinoma (grade I). The common way to estimate the epithelial area quantitatively is by interactive morphometry with point counting. Due to the fact that such a method is tedious, time-consuming, and still bears a certain subjective element, digital image processing has been applied to estimate the epithelial area automatically. A similar approach has been applied to ovarian tumors; the image processing results have been shown to be strongly correlated to the interactively assessed epithelial area percentages.

In this image processing study, the focus is on segmenting the image into a portion occupied by epithelium and a portion occupied by stroma. Thereafter, the ratio of the number of image points (pixels) in the epithelium segment and the pixels in the tissue segment (epithelium plus stroma) gives an estimate of the epithelial area percentage. The method for segmentation of the image is based on the observation that epithelial nuclei in ovarian tumors are more tightly packed than stromal nuclei. To a certain extent, this is also true for endometrial carcinomas. However, several factors may complicate the segmentation of the image, especially the tight packing of the stromal nuclei in endometrial tumors. In addition, cutting artefacts, such as tears and folds in the tissue, and variability in staining intensity complicate the distinction of epithelium and stroma.

This report describes the modified image processing method used to assess the epithelial area percentage in 30 endometrial carcinomas and compares it with the methods used for endometrial and ovarian cancers.
MATERIALS AND METHODS

Patient Material

In this study, 30 stage I endometrial carcinomas of varying histologic grades were used. Only hysterectomy specimens were studied. In all cases, myometrial invasion occurred, in 14 cases, less than one third and in 16 cases, more than one third of the thickness of the myometrial wall. Based on well-defined histopathologic criteria, the carcinomas were diagnosed by one of us (J.P.A.B.) as grade I (n = 10), grade II (n = 14) or grade III carcinoma (n = 6). Tissue was fixed in 10% neutral formalin (pH of 7.0) for approximately 24 hours at room temperature. It was then dehydrated in alcohol of increasing strength and embedded in paraplast (57°C). Sections 4 μm-thick were prepared and stained with hematoxylin-eosin for visual inspection.

To discriminate epithelium from stroma, it is important to emphasize that in the previously described image processing method, which has been proven to be successful in ovarian tumors, distinction of nuclei and cytoplasm is essential. For that purpose, the standard hematoxylin-eosin stain is not suitable, because of insufficient spectral specificity of the dyes and the absence of tissue-component specificity. Instead, the Feulgen and naphthol yellow combination stain, a stoichiometric, component-specific stain, gives good spectral separation of the dyes and has been used in this study to facilitate the image segmentation.

In the standard Feulgen procedure, sections are hydrolyzed in 5N HCl for 30 minutes and stained with a solution of 0.5 g pararosanilin in a mixture of 15 mL 1N HCl and a solution of 0.5 g K2S2O5 in 85 mL distilled water for 45 minutes. Sections then are stained for 5 seconds with a solution of 0.1 g naphthol yellow in a mixture of 100 mL distilled water and 1 mL acetic acid.

Image Acquisition

The stain caused the nuclei to turn reddish-brown and the remaining tissue components were stained yellow. A monochromatic blue filter (λ = 420 nm; Δλ < 10 nm; Schott, Tiel, Netherlands), for which naphthol yellow shows maximum absorption, was used to distinguish all tissue area from the surrounding background area. A monochromatic yellow filter (λ = 552 nm; Δλ < 10 nm), for which the Feulgen stain shows maximum absorption, was used to discriminate nuclei from the cytoplasm and lumina.

A blue-yellow image pair was recorded per field using the two filters sequentially on a Universal Engineering Microscope (Carl Zeiss, Oberkochen, Germany) connected to a Chalnicon TV-camera (Bosch, Stuttgart, Germany). A dry *6.3 objective with a numerical aperture of 0.16 (Carl Zeiss) was used. This magnification resulted in a pixel-to-pixel distance of 90 μm at the specimen level. The images were analyzed using an image processing system (Kontron Bildanalyse GmbH, Eching, Germany).

Interactive Morphometry

To evaluate the image processing results, control percentages of the epithelial and stromal areas were measured interactively. Such measurements, using point counting, have been shown to provide a good reference for the image processing results. The point counting method was performed on one blue-yellow image pair subjectively selected from the most epithelium-rich areas of the specimen. For assessment of the control percentage of epithelium, a 108-point regular grid was randomly positioned on the blue and yellow images and points overlying epithelium, stroma (together forming all tissue), and lumen were accumulated separately. In this study, the control percentage of epithelium for the endometrial carcinomas ranged from 51.8% to 82.1%.

Image Processing

The image processing method for the automated assessment of the epithelial percentage consists of two parts: processing the blue image to determine the area of tissue (epithelium and stroma) and processing the yellow image to estimate the epithelial area.

This two-part method is similar to the method used for ovarian tumors, with the exception of the tuning of some parameter values.

In quantitative microscopy, it is important to correct images for shading due to uneven illumination in the microscope and uneven recording effects in the camera. For this purpose, at the beginning of each day, a blank-field (i.e., on the slide but outside the tissue section) image pair was recorded and smoothed (refer to ref 8) by rank filtering (rank, 80%; window, 15*15 pixels), averaging, and rank filtering again (rank, 50%; 13*13 pixels). This blank-field image pair, free from noise and dust particles, was then used as a reference image pair. Each recorded image was always corrected for shading before further processing.

The image processing for the blue image consists of the steps described below (Fig 1).

Shading correction. The image is corrected for shading by subtracting the blue reference image.

Segmentation of tissue. The image is segmented with a well-known global threshold technique. The threshold is derived from the grey value histogram. If the smoothed histogram has only one peak, the image is assumed to contain tissue only. If the smoothed histogram has two peaks, the image is assumed to contain tissue and lumina. According to the Bayes rule, to minimize the probability of misclassifying an object point as a background point and vice versa, the minimum between the two peaks is selected for the threshold level.

Elimination of artefacts. After thresholding, unresolved issues are the small artefacts in the background area and the tears or holes in the tissue area. Before eliminating these artefacts, a closing operation (two dilations followed by two erosions) is applied to fill the small holes. Small artefacts, such as dirt, mucus, or loose cells in the thresholded binary image, are eliminated by a tenfold erosion (four connected, i.e., structure element is a cross); this is effective for objects smaller than 40 μm. This procedure is followed by a propagation, with the thresholded binary image as a mask, to return to the original dimensions. Small tears or holes in the tissue area are removed in the inverted binary image if the area is less than 100 pixels, i.e., 400 μm². After an inversion of the binary image, this results in the estimated area of tissue.

The yellow image of the same microscopic field is subsequently processed automatically to determine the area of epithelium. The image segmentation of the yellow area is based on the observation that epithelial nuclei are generally more tightly packed than stromal nuclei. The processing of the yellow image consists of the steps described below (Fig 1).

Shading correction. The image is corrected for shading by subtraction of the yellow reference image.

Gaussian blurring. Given that epithelial nuclei are
more tightly packed than stromal nuclei, segmentation of
the epithelium is facilitated by bridging the gaps between
epithelial nuclei (by the application of a 3×3 linear Gaussian
filter) without bridging the gaps between less closely packed
stromal nuclei.

Segmentation of epithelial area. In the yellow images, the
grey value distribution is found to be unimodal in all cases
but there are strong variations in the shapes of the different
histopathologic tumor types. Because the threshold selec-
tion algorithm has to be resistant against the variability in
staining intensity, the threshold selection is based on the
image contrast (i.e., the difference between the minimum
value in the histogram and the lumen modal value) (referred to as
min of the lumen modal value). The difference between the
minimum value and the lumen modal value is calculated as follows:

\[
\text{threshold level (L)} = 0.235 \times \log(\text{min modal})
\]

From the image contrast range (i.e., log(min) - log(mod L)) and the cytoplasm modal value (mod c), the threshold level (Ly) is computed as follows: log(Ly/Mod c) = 0.235 log(min/modal L).

Elimination of artefacts. Again, unresolved issues, such as
holes, stromal nuclei, and gaps between epithelial nuclei
in the thresholded binary image, remain at this point. To
prevent elimination of single epithelial nuclei, small holes
are filled using a closing operation (four connected) with a
step size of one pixel (i.e., one dilation followed by one ero-
sion) before small objects are eliminated. Small objects,
mostly stroma nuclei, are then eliminated by the application
of the skeleton operation followed by a closing (four con-
ected) with a step size of two pixels to fill small epithelial
gaps and the removal of skeleton objects not exceeding
50 pixels. The result is the so-called tentative epithelial area.

Small gaps between epithelial nuclei in the tentative epitel-
hal area are closed if the area is less than 200 pixels (800 µm²). Finally, the estimate for the segmented epithelial

From the image contrast range (i.e., log(min) - log(mod L)) and the cytoplasm modal value (mod c), the threshold level (Ly) is computed as follows: log(Ly/Mod c) = 0.235 log(min/modal L).

Elimination of artefacts. Again, unresolved issues, such as
holes, stromal nuclei, and gaps between epithelial nuclei
in the thresholded binary image, remain at this point. To
prevent elimination of single epithelial nuclei, small holes
are filled using a closing operation (four connected) with a
step size of one pixel (i.e., one dilation followed by one ero-
sion) before small objects are eliminated. Small objects,
mostly stroma nuclei, are then eliminated by the application
of the skeleton operation followed by a closing (four con-
ected) with a step size of two pixels to fill small epithelial
gaps and the removal of skeleton objects not exceeding
50 pixels. The result is the so-called tentative epithelial area.

Small gaps between epithelial nuclei in the tentative epitel-
hal area are closed if the area is less than 200 pixels (800 µm²). Finally, the estimate for the segmented epithelial
area is improved by a dilation (eight connected, i.e., structure element is a square). The result of this procedure is the area of epithelium, estimated from the yellow image.

The ratio of the number of pixels in the estimated area of epithelium and the estimated area of tissue gives an estimate of the percentage of epithelium in the image.

With this two-step automated image processing method, four image pairs per slide with a total area of 3.2 mm² at specimen level are recorded from the most epithelium-rich areas of the specimen. The maximum over these four image pairs is considered to represent the area percentage of epithelium of the specimen.

Differences Between the Image Processing Methods Applied to the Endometrium and Ovaries

The differences between the image processing methods for the endometrium and ovaries lie in the threshold selection method for the yellow image (step no. 3) and the elimination of artefacts in the binary blue and yellow image (steps no. 3b and 4a through d) (Fig 1).

In the threshold selection for the yellow image in the ovary, the minimum pixel value has been defined as either the 0.5% or 5.0% extreme minimum value, depending on the value of the slope between the two "minima" in the grey value histogram. If the slope exceeds a critical value of 25.0, darkly stained lymphocytes are present in the epithelium, and the 0.5% minimum pixel value will result in too low a threshold level. Therefore, the 0.5% is then replaced by the 5.0% minimum pixel value. In the endometrium, the slope always exceeds the above-mentioned critical value due to the presence of darkly stained nuclei and lymphocytes in the stroma. For that reason, the minimum is fixed at the 0.5% pixel value.

The operation, which in ovarian tumors fills holes in the binary image after thresholding and aims at bridging the gaps between epithelial nuclei, is omitted for endometrial carcinomas. In the endometrium, this filling operation falsely concatenates stromal nuclei, caused by a tighter packing of these nuclei compared with ovarian tumors.

Ovarian tumors, with the exception of the endometrioid type, generally show a smooth and broad morphologic pattern of stroma and lumen; in endometrioid carcinomas, this pattern is more curly and narrow. To close small holes in the tissue and epithelium in the ovary, skeleton, four, and eight connected erosion operations have to be applied. Use of skeleton and erosion operations in the endometrium, however, will fill narrow lumen and stromal areas. For that reason, only holes in endometrial tumors with an area not exceeding a certain critical value are filled, and the erosion operations are excluded. Due to a more curly pattern of the epithelium/stroma contours in endometrioid carcinomas, a dilation operation (eight connected) is used to get a more realistic and accurate estimation of the epithelial area. The method to close small holes in tissue and epithelium and the afore-mentioned eight connected dilation operation, both used in endometrioid carcinomas, can also be used in ovarian tumors without significantly affecting the estimated epithelial area. Due to a curly and narrow pattern, similar to the pattern in endometrial carcinomas, these two steps will affect the image processing result only in endometrioid ovarian tumors.

RESULTS

Evaluation of the image processing method consisted of two parts. First, the tissue segmentation results and the epithelium segmentation results were tested using a training set of ten cases of carcinoma with varying degrees of malignancy (two grade I cases, six grade II cases, and two grade III cases). A typical example of the tissue segmentation (resulting in the blue image) and the epithelium segmentation
(resulting in the yellow image) is illustrated in Fig 2A and B, respectively. In Fig 2, the white, two pixel-thick lines represent the contours of the estimated areas of tissue (left) and epithelium (right).

From a visual comparison of the processed images with the original images, it was concluded that the method produced satisfactory segmentation results. The estimated epithelial percentage ranged from 64% to 88% for the specimens in the training set. From the visual comparison, it appeared that for only one of the six grade II carcinomas the area percentage of epithelium was slightly overestimated, due to darkly stained and closely packed stromal nuclei.

In the second part of the evaluation, results of the image processing method were obtained from all 30 stage I endometrial carcinomas, including the carcinomas of the training set. Visual inspection of the segmented images led to the conclusion that the method produced reliable results. The estimated area percentage of epithelium varied from 60% to 93%. An overestimation of the epithelial percentage appeared for seven of the 30 specimens (four grade I, two grade II, and one grade III carcinomas), caused by darkly stained and tightly packed stromal nuclei. For the moderately and poorly differentiated carcinomas, this overestimation was negligible compared with the overall result for the specimen. Figure 3 shows two examples of the epithelium segmentation, resulting in the yellow image. The white, two pixel-thick lines represent the contours of the estimated areas of epithelium. From Fig 3, it can be seen that a few small stromal areas are closed and are considered part of the epithelium.

Our method of estimating the epithelial percent age of the specimen is based on the analysis of four microscopic fields selected from the most epithelium-rich areas. The selection of these microscopic fields may affect the outcome of the image processing method, as the amount of epithelium is not uniformly distributed throughout the slide. To evaluate this influence on the estimated epithelial percentage, the image processing procedure was applied a second time to 15 endometrial carcinomas with varying degrees of malignancy. The results were compared with the epithelial percentages estimated in the first analysis, as shown in Fig 4. The correlation between the first and second assessments of the epithelial percentage is 0.980 ($P < .001$; range of the linear regression coefficient, 0.90 to 0.94). From this result, it can be concluded that the reproducibility and consistency under repeated field selection are adequate.

To evaluate the image processing method, the estimated epithelial percentages were compared with the control percentages, which resulted from point counting (Fig 5). From Fig 5, it can be seen that the epithelial percentages are strongly correlated to the control percentages (correlation coefficient, $r = .987$, $P < .001$; range of the linear regression coefficient, 1.00 to 1.02).

The differences between the image processing methods for ovary and endometrium, as described in the Materials and Methods section, resulted in different tissue and epithelium segmentations. Figure 6 illustrates an example of a blue-yellow image pair recorded from an endometrial carcinoma. Comparison of the tissue segmentations, resulting from the ovarian method (Fig 6, top left) and the endometrial method (Fig 6, top right), shows that small lumen

**FIGURE 3.** Examples of yellow images (×63 objective) from endometrial carcinomas stained with pararosanilin-Feulgen and naphthol yellow. The white two pixel-thick lines give the contours of the estimated areas of epithelium (left and right).
areas, which are closed in Fig 6, top left, and counted as part of the tissue area, remain part of the lumina in Fig 6, top right. Comparison of the epithelium segmentations (Fig 6, bottom left and bottom right) shows that stromal areas, which are counted as part of the epithelium in the ovarian method (Fig 6, bottom left), remain part of the stroma in the endometrial method (Fig 6, bottom right).

DISCUSSION

The objective of this study was to investigate the possibility of automatically estimating the area percentage of epithelium in endometrial cancer sections. The image processing method has been based on the observation that epithelial nuclei are generally more tightly packed than stromal nuclei. Estimates of the epithelial percentage have been obtained for 30 cases of stage I endometrial carcinomas of varying histologic grades. Despite the observation that the differences in packing between epithelial and stromal nuclei were less than those in ovarian tumors, it can be concluded from visual inspection that the tissue and epithelium segmentation results are satisfactory. For seven specimens in this study, there was an overestimation of the epithelial percentage due to darkly stained and tightly packed stromal nuclei. For the moderately and poorly differentiated carcinomas, this overestimation was found to be negligible compared with the overall results for the specimen.

The strong correlation between the image processing results and the control percentages, resulting from interactive morphometry, indicates that this method can be used to estimate the epithelial percentage in endometrial cancer sections. Image processing offers some practical advantages over the interactive morphometrical assessment: the automated method is less time-consuming and more reproducible.

The remaining interactive element in the image processing method is the selection of microscopic fields from the most epithelium-rich areas of the specimen, which may affect the estimated epithelial area. Although the potential influence of the field selection has been shown to be negligible, this selection may be automated by the application of a simplified image processing method at a lower magnification (objective, ×2.5). This method is currently being developed.

The image processing method for endometrial carcinomas is based on the method developed for ovarian tumors, described elsewhere in detail. In the ovarian model, it is assumed that the epithelial nuclei are more tightly packed and more darkly stained than the stromal nuclei. This assumption is not always true for endometrial tumors due to a darker staining and tighter packing of the stromal nuclei. A tight packing of the stromal nuclei can produce epithelial areas in the stroma, which results in an overestimation of the epithelial percentage. This problem could be overcome by tuning some parameters in the image processing method, as described in the Materials and

![Figure 4](image-url)  
**Figure 4.** The first and second assessments of the epithelial area percentage by digital image processing for 15 endometrial specimens with varying degrees of malignancy. The correlation coefficient, \( r \), is 0.90 and the linear regression coefficient ranges between 0.90 and 0.94.

![Figure 5](image-url)  
**Figure 5.** Results of the image processing method, %E/E + S, related to control percentages, %Ea/Ea + Sa. The correlation coefficient, \( r \), is 0.987 and the linear regression coefficient ranges between 1.00 and 1.02. "E" indicates the number of pixels in the epithelial portion and "S" indicates the number of pixels in the stromal part, measured by digital image processing. "Ea" indicates the number of points overlying the epithelial portion and "Sa" indicates the number of points overlying the stromal part, established morphometrically.
Methods section. The above-mentioned assumption, however, is not valid for endometrial hyperplasias. Because of this, the method results in an overestimation of the epithelial area and is therefore not applicable to endometrial hyperplasias. A second assumption in the ovarian model is the smoothness of the epithelium/humen contour and a broad pattern of the lumina and stromal areas. This observation is not valid for endometrial carcinomas, which show a more curly and narrow pattern. The operations used in the ovary, eliminating small holes in the tissue and the epithelium, will result in classifying small lumina and stromal areas as epithelial areas, which gives an overestimation of the epithelial percentage in endometrial carcinomas. For that reason, a different approach has been applied in endometrial carcinomas, which takes into account the size of the hole. Such an approach will also be applicable in ovarian tumors.

It can be concluded that the greater part of the image processing method, automatically estimating
the area percentages of epithelium and stroma, is equivalent for both ovary and endometrium and yields reliable results. Therefore, it may be expected that this method is also applicable to tumors of other organs, such as prostatic cancers, breast cancers, and malignant melanomas.

REFERENCES

8. Schipper NW, Smeulders AWM, Baak JPA: Quantification of epithelial volume by image processing applied to ovarian tumors. Cytometry 8:345-352, 1987