



**UvA-DARE (Digital Academic Repository)**

**Automated estimation of epithelial volume in breast cancer sections. A comparison with the image processing steps applied to gynaecologic tumors**

Schipper, N.W.; Smeulders, A.W.M.; Baak, J.P.A.

*Published in:*

Pathology, Research and Practice

*DOI:*

[10.1016/S0344-0338\(11\)80264-5](https://doi.org/10.1016/S0344-0338(11)80264-5)

[Link to publication](#)

*Citation for published version (APA):*

Schipper, N. W., Smeulders, A. W. M., & Baak, J. P. A. (1990). Automated estimation of epithelial volume in breast cancer sections. A comparison with the image processing steps applied to gynaecologic tumors. *Pathology, Research and Practice*, 186, 737-744. DOI: 10.1016/S0344-0338(11)80264-5

**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.

## Automated Estimation of Epithelial Volume in Breast Cancer Sections\*

### A Comparison with the Image Processing Steps Applied to Gynecologic Tumors

---

N. W. Schipper, A. W. M. Smeulders and J. P. A. Baak

Department of Pathology, Free University, Amsterdam (N.W.S., J.P.A.B.), and Faculty of Mathematics and Computer Science and Faculty of Biology, University of Amsterdam, Amsterdam (A.W.M.S.), The Netherlands

#### SUMMARY

*The paper describes an image analysis technique for automated estimation of the epithelial percentage in standard paraffin tissue sections of invasive ductal breast cancers. Two staining procedures are evaluated: Feulgen (pararosanilin) and CAM 5.2-demonstrating the presence of cytokeratin 8 and 18-, both counterstained with naphthol yellow.*

*In the technique, one image is recorded with a filter to visualize where the epithelium lies. This filter is chosen corresponding to the type of staining: it is yellow for Feulgen and blue for anti-cytokeratin CAM 5.2. To visualize where the stroma lies, the same image can be used for anti-cytokeratin CAM 5.2, whereas for Feulgen, a second image has to be recorded from the same microscope field with a blue filter. The image processing steps to determine the total tissue area comprise correction for shading, segmentation of the tissue area, and restoration of the segmented image by removal of small artefacts and closure of small tears in the tissue. The method for determination of the epithelial area consists of the following steps: correction for shading, gaussian blurring, segmentation of nuclei or epithelial cells, and editing of the segmented image by removal of small objects and closure of small spaces between the epithelial nuclei or cells. These image processing steps are compared to those for quantification of the epithelial percentage in gynecologic tumors of epithelial origin.*

*For the Feulgen stain, the method is evaluated on 30 breast cancers of the ductal type (4 grade I, 12 grade II, and 14 grade III). Comparison with the control percentage, established interactively through point-counting, shows a strong correlation for all 30 specimens ( $r = 0.98$ ). For the anti-cytokeratin stain CAM 5.2, results of the method are obtained with 28 breast cancers of the ductal type (4 grade I, 11 grade II, and 13 grade III), of which 18 specimens are overlapping with the Feulgen specimens. The estimated epithelial percentage, resulting from the Feulgen stained specimens, shows a strong correlation with the image processing result obtained with the corresponding CAM 5.2 stained specimen ( $r = 0.99$ ).*

*It can be concluded that the described image analysis technique results in a reliable estimate of the percentage of epithelium, and may be used to select the epithelium-rich areas automatically.*

\* This study was supported in part by grant 28-736 of the Praeventiefonds, and the SPIN programs 3 DAI and 3 DCV.

## Introduction

Breast cancer is the primary cause of death with a mortality rate of 30 per 100 000 inhabitants in western Europe. In spite of many developments in hormonal and (adjuvant) chemotherapy, or surgical treatment, prognosis has only slightly improved over the past decades<sup>5</sup>. Tumor size and nodal status have been widely used to predict the prognosis. However, their predictive value is not very high<sup>6</sup>, which may result in undertreatment or overtreatment<sup>13</sup>. Although nuclear and histological grade have prognostic implications and thus are of some use, assignment is subjective and not well reproducible<sup>21,22</sup>. Therefore, more objective, reproducible and stronger prognostic features are urged for, describing morphology and behaviour of the tumor itself.

Nuclear DNA content and morphometrical features such as nuclear size, cellularity index, and the number of mitoses are examples of such features. They have been shown to be accurate prognosticators in breast cancer<sup>1,2,4,11</sup>, including the epithelial percentage. As to the morphometrical features, the results are greatly depending on the selection of the microscope field(s)<sup>9</sup>. To investigate the (prognostic) value of the epithelial percentage, and to select the most cellular areas automatically, it is useful to estimate the volume percentage of epithelium in breast cancers.

The relative volume of epithelium is estimated through measurement of the area percentage of tissue in a microscope field covered by epithelium. Digital image processing has been shown to measure the epithelial area reliably in gynecologic tumor sections<sup>16-19</sup>. In these studies, two staining procedures have been evaluated: Feulgen (pararosanilin) and anti-cytokeratin CAM 5.2, both counterstained with naphthol yellow.

The paper describes an image analysis technique for automated estimation of the epithelial percentage in breast cancer sections of the ductal type. Results of the method are obtained with 30 Feulgen stained specimens and 28 anti-cytokeratin CAM 5.2 stained specimens of varying histological grades. For both combination stains, the image processing method is compared with the corresponding methods for ovarian tumors and endometrial hyperplasias and carcinomas.

## Material and Methods

### Staining Methods

In the Feulgen-naphthol yellow part of the study 30 breast cancers of the ductal type and with varying histological grades have been analyzed. Based on well defined histopathological criteria, of the breast cancers, 4 were diagnosed as grade I, 12 as grade II and 14 as grade III.

Tissue was fixed in 10% neutral formalin, i.e.,  $6.5 < \text{pH} < 7.5$ , for approximately 24 h at room temperature, dehydrated in alcohol of increasing strength, and embedded in paraplast at 57°C. Sections, 4  $\mu\text{m}$ , were prepared and stained with Feulgen (pararosanilin) and naphthol yellow. In the Feulgen procedure, sections are hydrolyzed in 5 N HCl for 30 minutes and stained with a solution of 0.5 g pararosanilin in a mixture of 15 ml 1 N HCl and a solution of 0.5 g  $\text{K}_2\text{S}_2\text{O}_5$  in 85 ml distilled water for

45 minutes. Then, sections are counterstained 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.

In the anti-cytokeratin CAM 5.2-naphthol yellow part of the study 28 breast cancers of the ductal type and with varying histological grades have been analyzed. Of these specimens, 4 were diagnosed as grade I, 11 as grade II and 13 as grade III; 3 grade I, 7 grade II, and 8 grade III of these breast cancers are overlapping with the patient material used in the first part of the study.

Tissue was fixed in 10% neutral formalin, i.e.,  $6.5 < \text{pH} < 7.5$ , for approximately 24 h at room temperature, and embedded in paraplast at 57°C. Paraplast sections were cut at 4  $\mu\text{m}$  and mounted on slides, which were pretreated with 0.1% Poly-L-Lysine (Sigma, St. Louis, Missouri) and dried overnight at 37°C. Sections were dewaxed, rehydrated to water and digested with 0.5% trypsin (Flow Labs, UK) in 0.5%  $\text{CaCl}_2$ , with pH 7.8, for 30 minutes at 37°C. In order to block endogenous peroxidase, sections were placed in a mixture of 0.3%  $\text{H}_2\text{O}_2$  in methanol for 30 minutes. After thorough washing in phosphate-buffered saline (PBS, pH 7.4), sections were incubated with normal rabbit serum, diluted 1:50 in PBS containing 1% bovine serum albumin for ten minutes. Thereafter, sections were incubated for 18 h with CAM 5.2 diluted 1:10, monoclonal antibody raised in mice against cytokeratins 8 and 18 (Becton & Dickinson, Sunnyvale, California, USA) at 4°C<sup>10</sup>. After a second thorough washing in PBS, sections were incubated for 30 minutes with biotin-labeled rabbit immunoglobulins to mouse immunoglobulins (Dakopatts, Copenhagen, Denmark) at a dilution of 1:400 in PBS-BSA. Then sections were washed in PBS and incubated with avidin-biotin HRP-conjugated serum (Dakopatts, Denmark) with dilution 1:100 for 60 minutes. The peroxidase was made visible by immersion in a mixture of 0.5% 3,3'-diaminobenzidine, 0.02%  $\text{H}_2\text{O}_2$  in PBS, and a solution of 2 ml 1%  $\text{NiCl}_2$  in distilled water per 100 ml, for five minutes. Hereafter, sections were rinsed in water, counterstained for 3 to 4 minutes with a solution of 0.1 g naphthol yellow in a mixture of 100 ml distilled water and 1 ml concentrated acetic acid, dehydrated in alcohol, cleared in xylol, and mounted in Depex.

### Equipment

The Feulgen stain colors the nuclei reddish-brown and the naphthol yellow stain colors the remaining tissue components yellow. A monochromatic blue filter with  $\lambda = 420 \text{ nm}$ , and width  $\Delta\lambda < 10 \text{ nm}$  (Schott, Tiel, The Netherlands), for which naphthol yellow shows maximum absorption, is used to distinguish all tissue area from the surrounding background area. A monochromatic yellow filter with  $\lambda = 552 \text{ nm}$  ( $\Delta\lambda < 10 \text{ nm}$ ), for which Feulgen pararosanilin shows maximum absorption, is used to discriminate nuclei from the cytoplasm and the background.

The CAM 5.2 stain colors the epithelial cells brown and naphthol yellow colors the remaining tissue components yellow. A monochromatic blue filter with  $\lambda = 470 \text{ nm}$  ( $\Delta\lambda < 10 \text{ nm}$ ), for which anti-cytokeratin CAM 5.2 shows maximum absorption, is used to distinguish epithelium from stroma, and tissue from background. When it is concluded that the keratin staining result is too weak, a second image is recorded with a monochromatic yellow filter with  $\lambda = 576 \text{ nm}$  ( $\Delta\lambda < 10 \text{ nm}$ ) to better distinguish epithelium from stroma.

The images are recorded on a Zeiss UEM microscope (Carl Zeiss, Oberkochen, FRG) connected with a Chalmicon TV-camera (Bosch, Stuttgart, FRG). A dry 6.3\* objective with a N.A. of 0.16 (Carl Zeiss, Oberkochen, FRG), corresponding to a pixel-to-pixel distance of 2.0  $\mu\text{m}$  at specimen level, is used. The images are analyzed on a Kontron Image Processing System (Kontron Bildanalyse GmbH, Eching, FRG).

### Interactive Morphometry

To assess the quality of the image processing method, control percentages of the epithelial and stromal areas are measured interactively through point-counting. For assessment of the control percentage, a 168-point regular grid<sup>23</sup> is positioned on the blue and yellow images and points overlying epithelium, stroma (together forming all tissue) and lumen are accumulated separately. In this study, the control percentages of epithelium for the Feulgen stained breast cancer sections range from 25.0% to 92.2%.

### Prognostic Evaluation

To study the prognostic value of the epithelial percentage, estimated by image processing, in ductal breast cancers, univariate survival analysis according to Kaplan-Meier is performed<sup>8</sup>. Differences between survival curves are analyzed using the Mantel-Cox test<sup>12</sup>, with  $p < 0.05$  considered significant. The analysis has been performed with the aid of the BMDP package, using the program Life Tables<sup>3</sup>.

### Image Processing

Two methods are described here: one for Feulgen images, named F-method, and another one for anti-cytokeratin CAM 5.2 stained sections (C-method). Both methods are similar to the corresponding ones for ovarian tumors<sup>17,19</sup> and endometrial carcinomas and hyperplasias<sup>18,19</sup>, except for the tuning of some parameter values. In this section the image processing steps for both staining methods are described in brief. For a more detailed description see<sup>17-19</sup>.

Prior to further processing, each recorded image is always corrected for shading, due to uneven illumination and uneven recording effects. The method to correct images for shading is described in detail in<sup>17</sup>.

### F-Method

In this part, the image processing steps for Feulgen stained images, named F-method, are described.

The processing of the blue image, subsequent to digitization and correction for shading, consists of the following steps:

- 1) The image is segmented with a threshold technique<sup>7,15</sup>, where the threshold level is automatically derived from the grey value histogram through the mode-method<sup>14</sup>.
- 2) After thresholding, a closing operation is applied to fill small holes. Small artefacts, such as dirt, mucus or loose cells in the thresholded binary image are eliminated by a tenfold erosion (4-connected), effective for objects with a cross-size smaller than 40  $\mu\text{m}$ . Small tears or holes in the tissue area are removed if the area is less than 100 pixels, i.e., 400  $\mu\text{m}^2$ . This concludes the estimation of the tissue area from the blue image.
- 3) The processing of the yellow image starts from the observation that epithelial nuclei are more tightly packed than stromal nuclei. So, segmentation of the epithelium is facilitated by bridging the gaps between epithelial nuclei, by application of a 3\*3 linear Gaussian filter.
- 4) Also the yellow image is segmented with a global threshold, automatically derived from the grey value histogram. As the threshold selection algorithm has to be resistant against variability in staining intensity, the selection of the threshold level (lvl) is based on the image contrast (i.e.,  $\log(\text{min}) - \log(\text{mod}_1)$ ) and location of the highest tissue peak value ( $\text{mod}_c$ ), see<sup>12</sup>. The threshold level is computed from  $\log(\text{lvl}/\text{mod}_c) = 0.235 \times \log(\text{min}/\text{mod}_1)$ . The minimum pixel value (min) is defined as the grey value skipping either 0.5% or 5.0% of the extreme pixel

values, depending on the slope between these two minimum values in the grey value histogram.

5) First, small holes are filled using a closing operation (4-connected) with a stepsize of 1 pixel to prevent elimination of single epithelial nuclei. Then, small objects, mostly stromal nuclei, are eliminated (by application of the skeleton operation, followed by a closing (4-connected) with a stepsize of 2 pixels, and removal of skeleton parts not exceeding 20 pixels). Small gaps between epithelial nuclei are closed if the area is less than 200 pixels (800  $\mu\text{m}^2$ ). Finally, the estimate for the segmented epithelial area is corrected by a dilation (8-connected).

### Differences Among Tumors of the Ovary, Endometrium and Breast

The visual differences in tissue architecture among tumors of the ovary, endometrium, and breast concern the packing of stromal cells, the pattern of epithelial structures, and the size and shape of the different components in the image. The image processing steps for automated estimation of the epithelial percentage are related to these characteristic differences. As a result, the F-methods differ in the threshold selection method for the yellow image and the elimination of artefacts in the binary blue and yellow images.

The potential presence of darkly stained lymphocytes in the epithelium in combination with a moderate to large difference in packing between epithelial and stromal cells, as is the case for epithelial ovarian tumors and breast cancers, affects the contrast range in the image. As the threshold selection for the yellow image should be resistant against this phenomenon, 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 these two minimum values in the grey value histogram. If the slope exceeds a predefined critical value, it is assumed that 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. Endometrial carcinomas are characterized by a relatively small difference in packing between epithelial and stromal cells and the appearance of leucocytoid cells in the stroma. Although these cells affect the contrast range as well, in combination with the resulting tighter packing of the stromal cells, the minimum has to remain fixed at the 0.5% pixel value. In these images, the alternative 5.0% minimum pixel value will result in too high a threshold level, which gives candidate epithelial areas in the stroma.

Ovarian tumors of the common epithelial types are characterized by a relatively broad pattern of stromal and luminal areas, whereas this pattern is more narrow in endometrial carcinomas and high grade breast cancers. As a result, the filling operation after thresholding which aims at bridging the gaps between epithelial nuclei in ovarian tumors, is omitted for both endometrial carcinomas and breast cancers. In both the endometrium and the breast, this filling operation falsely closes narrow stromal stalks.

The characteristic low cellularity of stroma in breast cancer sections, in combination with the presence of small epithelial parts in low grade breast cancers, affects the critical value for removal of stromal skeleton parts. This value is set at 20 pixels, and is chosen such to prevent elimination of small epithelial parts present in grade I and grade II breast cancers. In both the ovary and the endometrium, the stromal cells are more tightly packed. As a consequence, the value for elimination of stromal cells is set at 50 pixels.

Ovarian epithelial tumors are characterized by a relatively broad pattern of stroma and lumina. This characteristic enables the use of skeleton- and erosion operations for closing small holes

in the tissue and the epithelium. In high grade breast cancers and endometrial carcinomas, however, a more narrow pattern is noteworthy. So, small holes are closed if the area is smaller than a

certain critical value. Application of erosion operations will fill the characteristic narrow luminal and stromal stalks in these images. Due to a more curly pattern of the epithelium contour in

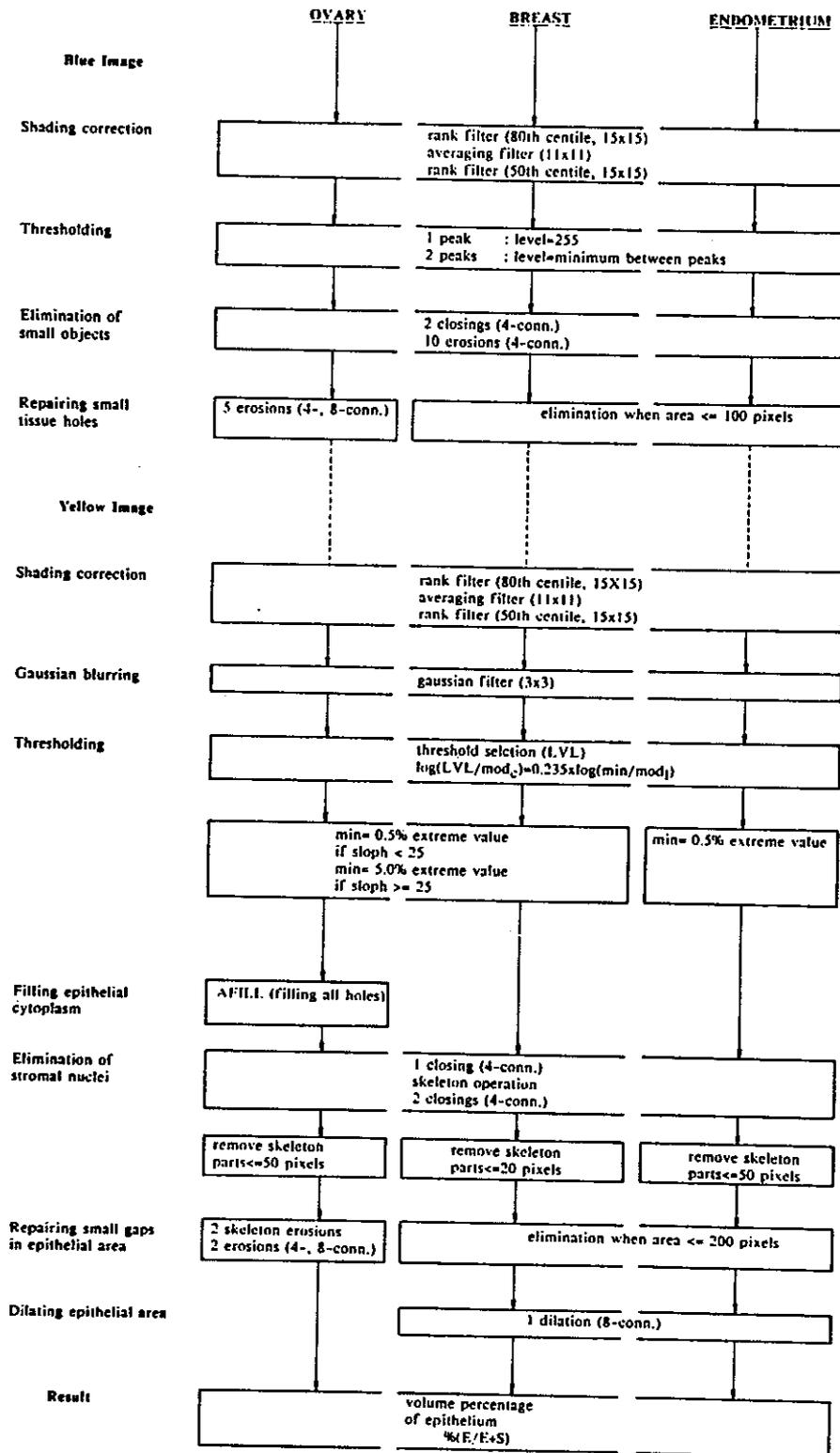


Fig. 1. Schematic representation of the different image processing steps for automated estimation of the epithelial percentage.

endometrial carcinomas and breast cancers, a dilation operation (8-connected) is used to get a reliable estimate for the epithelial area.

A schematic representation of the different image processing steps, applied to cancers of the ovary, endometrium, and breast, is given in Fig. 1.

### C-Method

The image processing steps for CAM 5.2 stained sections – demonstrating the presence of cytokeratin 8 and 18 –, named C-method, are described in this part.

The processing of the blue image, subsequent to digitization and correction for shading, consists of the following steps:

1. The image  $f(x, y)$  is segmented with one global threshold  $t_1$  to determine the total tissue area and a second one  $t_2$  to determine the epithelial area. The selection of both thresholds is based on the peaks or shoulders, denoted by  $f_1, f_2, f_3, f_4$ , and the valleys between them, named  $f_{12}, f_{23}, f_{34}$ , in the smoothed grey value histogram. It is assumed that the modal epithelial value never exceeds the value of  $C_c = 120$  and that the modal lumen value and the modal mucus value both always exceed the value of  $C_m = 180$ .

In the case that  $f_1$  is the lowest modal value exceeding  $C_m$ ,  $t_1 = f_{(1)}$  is used to threshold the tissue. When there is no modal value exceeding  $C_m$ , it is assumed that there is no lumen present and all image pixels are assigned to tissue.

In the case that  $f_1$  is the lowest modal value exceeding  $C_c$ ,  $t_2 = f_{(1)}$  is selected as the threshold to find epithelium. When all modal values in the histogram exceed  $C_c$ , either there is no epithelium present in the image or the cytokeratin stain is too weak to distinguish epithelium from stroma. Then, the yellow image  $g(x, y)$  is recorded and  $t_2 = g_{12}$  in the bimodal histogram is selected to threshold epithelium. When the histogram of the yellow image is unimodal and the modal value  $g_1 > C_m$ , it is assumed that epithelium is absent in the image; when  $g_1 < C_m$ , all image pixels are assigned to epithelium, see<sup>19</sup> for a more detailed description.

2. After tissue segmentation, remaining difficulties are small artefacts in the background area and tears or holes in the tissue. First, minimal tissue holes are filled using a closing operation (4-connected) with a stepsize of 2 pixels, filling holes with a cross-size smaller than 8  $\mu\text{m}$ . Then, small objects in the background are eliminated by application of a tenfold erosion operation (4-connected), which removes objects with a cross-size smaller than 40  $\mu\text{m}$ , eliminating loose single cells, dirt, or staining material. Finally, small tears or holes in the tissue are closed if the area is less than 100 pixels (400  $\mu\text{m}^2$ ). This results in the estimated area of tissue.

3. Again unresolved issues remain after segmentation of the epithelium. They are dirt or staining material present in the stromal areas and gaps in the epithelium. To prevent elimination of single epithelial cells and to fit small epithelial gaps, first small holes in the binary image are filled by application of a closing operation (4-connected) with a stepsize of 2 pixels. Then, objects with an area less than 30 pixels (120  $\mu\text{m}^2$ ) are eliminated. Finally, small spaces between the epithelial cells are closed if the area is less than 30 pixels (120  $\mu\text{m}^2$ ). The result of this procedure is the estimated area of epithelium.

### Differences among Tumors of the Ovary, Endometrium, and Breast

The visual differences in tissue architecture among anticytokeratin stained tumor sections of the ovary, endometrium, and breast concern the size and shape of the different components in the image. As a result, the C-methods only differ in the elimination of

artefacts after segmentation of the tissue area and the epithelial area.

Breast cancer sections are characterized by relatively small epithelial areas in grade I cancers, and small stromal areas in grade II and III cancers. As a consequence, to prevent removal of these small epithelial and stromal regions, the critical value for the area to eliminate small objects and to close small spaces is set at 30 pixels. In both ovarian and endometrial specimens, small objects are eliminated if the area is less than 100 pixels, and small spaces between the epithelial cells are closed if the area is less than 150 pixels.

With the automated image analysis technique, per slide four images with a total area of 4.2  $\text{mm}^2$  at specimen level are recorded from the most epithelium rich areas of the specimen. The maximum over these four images is taken to represent the volume percentage of epithelium of the specimen.

### Results

Results of the image analysis technique are obtained for 30 breast cancers, Feulgen pararosanilin and naphthol yellow stained, and 28 breast cancers, anti-cytokeratin CAM 5.2 and naphthol yellow stained, of the ductal type with varying histological grades.

#### Results of the F-Method

From visual inspection of the resulting images, it can be concluded that the tissue and epithelium segmentations give good results. The estimated epithelial percentage ranges from 21% for a grade I cancer to 98% for a grade III cancer. An example of the tissue segmentation result in the blue image and the epithelium segmentation result in the yellow image is given in Fig. 2a and 2b, respectively. In this figure the white lines represent the contours of the estimated areas of tissue (a) and epithelium (b).

To evaluate the image processing method, the estimated epithelial percentages are compared with the control percentages, resulting from point-counting, for all 30 specimens used in this study, see Fig. 3. From this figure it can be seen that the epithelial percentages are strongly correlated to the control percentages. The correlation coefficient is 0.98 ( $p < 0.001$ ), and the linear regression coefficient ranges from 0.99 to 1.02.

The epithelial percentage, as established with digital image processing, in ductal breast cancers, appears as a strong prognosticator in univariate survival analysis. When this percentage exceeds the critical value of 65%, the prognosis is considerably worse than for patients with an estimated epithelial percentage of smaller than 65%, see Fig. 4.

#### Results of the C-Method

From visual inspection, it can be concluded that the tissue and epithelium segmentation results are adequate. The estimated epithelial percentages range from 15% for a grade I cancer to 97% for a grade III cancer. For all specimens used in this part of the study, the method results in a reliable estimation of the epithelial percentage. An

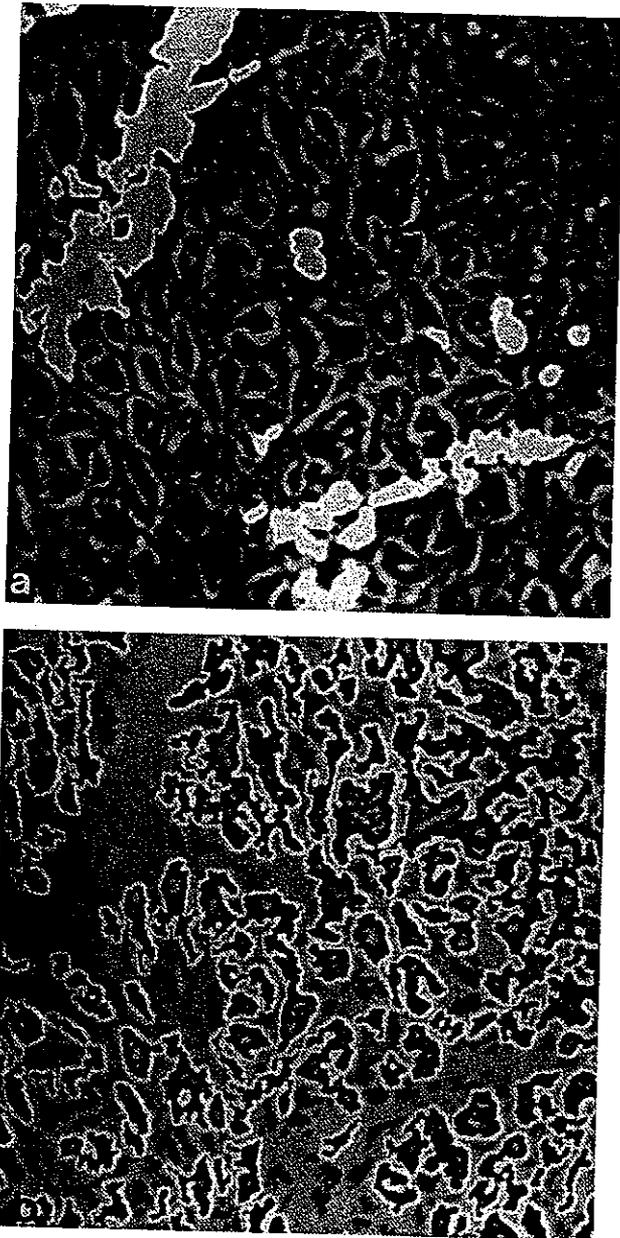


Fig. 2. Example of a blue and yellow image, coming from a breast cancer, stained with pararosanilin Feulgen and naphthol yellow. The white lines represent the contours of the estimated areas of tissue (a) and epithelium (b).

example of the epithelium segmentation result for a grade I cancer and a grade III cancer is given in Fig. 5a and 5b, respectively. Again, the white lines in this figure represent the contours of the estimated areas of epithelium.

For comparison of the image processing results with the epithelial percentages, resulting from image processing on Feulgen stained sections, 18 breast cancers with varying histological grades (grade I, n = 3; grade II, n = 7; grade III, n = 8) have been used. The results of the Feulgen and

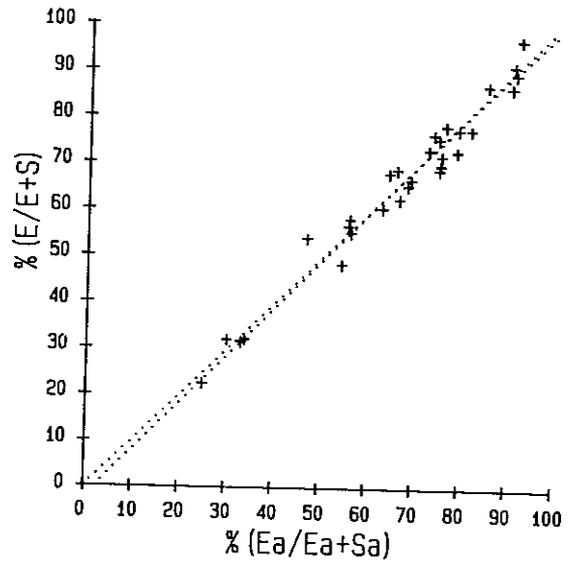


Fig. 3. Results of the image processing method,  $\% (E/E+S)$ , related to control percentages,  $\% (E_a/E_a+S_a)$ . The correlation coefficient is 0.98 and the linear regression coefficient ranges from 0.99 to 1.02. E is the number of pixels in the epithelial part and S the number of pixels in the stromal part, measured by digital image processing.  $E_a$  is the number of points overlying the epithelial part and  $S_a$  the number of points overlying the stromal part, established morphometrically.

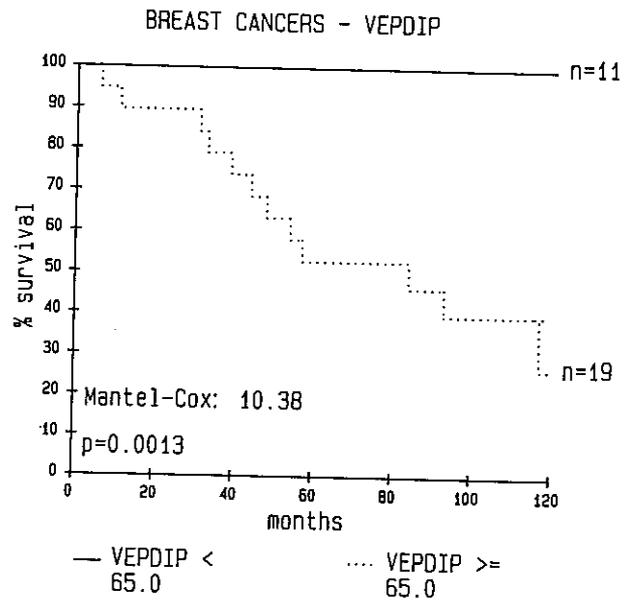


Fig. 4. Kaplan-Meier survival curves for patients with an estimated epithelial percentage (VEPDIP) of  $< 65\%$  (n = 11) and patients with a percentage  $\geq 65\%$  (n = 19).

CAM 5.2 image processing method are shown in Fig. 6. The correlation between the epithelial percentages is 0.99 ( $p < 0.001$ ), and the linear regression coefficient ranges from 0.87 to 0.90.

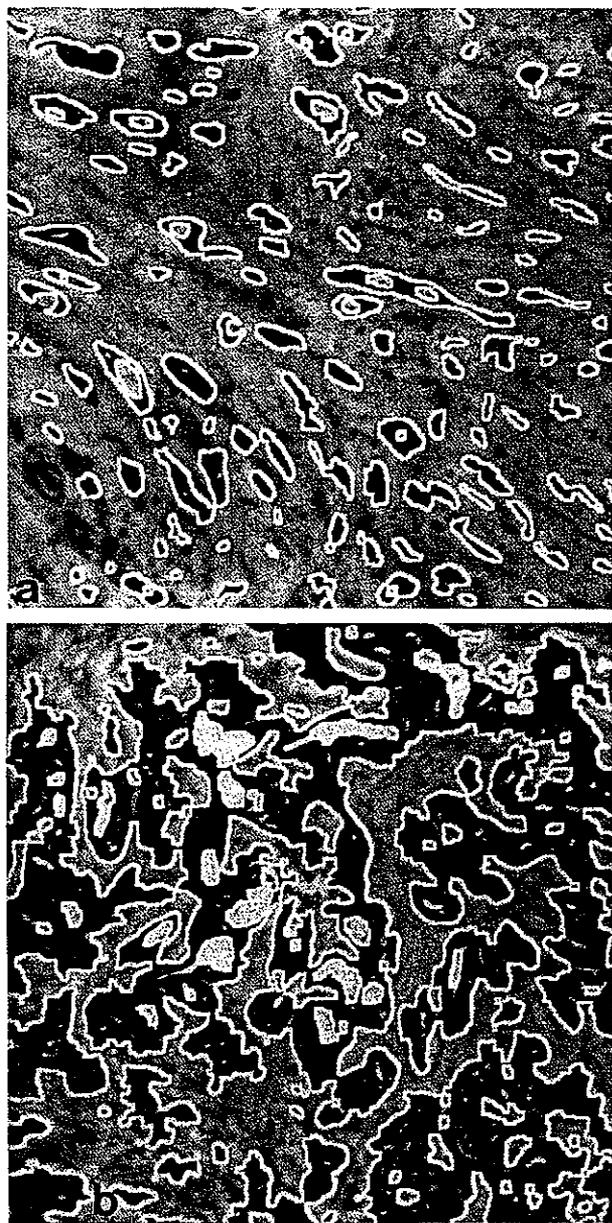


Fig. 5. Examples of blue images, coming from breast cancer sections, stained with anti-cytokeratin CAM 5.2 and naphthol yellow. The white lines represent the contours of the estimated areas of epithelium. — a: Image from a grade I breast cancer — b: Image from a grade III breast cancer.

### Discussion

The object of this study was to investigate the possibility of automatically estimating the volume percentage of epithelium in order to select the most cellular or epithelium-rich areas in breast cancer sections. The method was based on the previously developed image processing methods for estimation of the epithelial percentage in tumor sections of the ovary and endometrium, and was

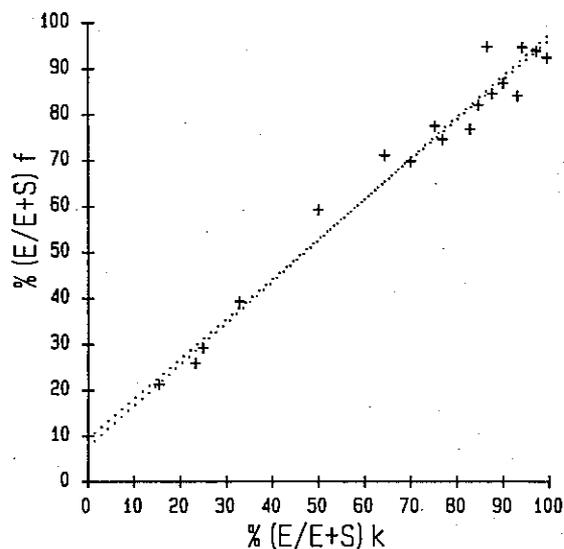


Fig. 6. Results of the Feulgen image processing method,  $\%(E/E+S)_f$ , related to the epithelial percentages obtained with CAM 5.2 stained specimens,  $\%(E/E+S)_k$ . The correlation coefficient is 0.99 and the linear regression coefficient ranges from 0.87 to 0.90.

implemented for two different staining procedures: Feulgen pararosanilin and anti-cytokeratin CAM 5.2, both counterstained with naphthol yellow.

Estimates of the epithelial percentage were obtained for 30 Feulgen stained and 28 CAM 5.2 (anti-cytokeratin 8 and 18) stained breast cancer sections of the ductal type, with varying histological grades. From visual inspection, it was concluded that the Feulgen method and the anti-cytokeratin method give satisfactory and reliable segmentation results for tissue and epithelium. The 'Feulgen' results were strongly correlated with the control percentages, resulting from interactive morphometry, for all 30 breast cancer sections ( $r = 0.98$ , see Fig. 3).

The epithelial percentage, as established with digital image processing of Feulgen stained breast cancer sections, is a strong prognosticator. When the percentage exceeds the value of 65%, the prognosis is worse than for patients with an estimated percentage smaller than 65%.

As to the biological explanation for the prognostic value of the epithelial percentage, it is understandable that the probability on distant metastases increases with an increase in the number of epithelial cells, and thus with the percentage of epithelium. This is the more important since the prognostic value of several proliferation markers is now well established, but adequate metastatic potential markers in breast cancer are lacking. Therefore, further analysis of the epithelial percentage as a feature correlated with the occurrence of metastases is desirable.

Comparison of the 'Feulgen' and 'anti-cytokeratin CAM 5.2' epithelial percentages, for 18 breast cancer sections of varying histological grades, showed a strong correlation ( $r = 0.99$ ).

The results were obtained by interactive selection of the microscope fields from the most epithelium-rich areas of

the specimen. It was demonstrated that this selection can be automated by application of a quicker, simplified version of the image processing methods at a lower resolution<sup>20</sup>. These simplified image processing methods, developed for tumors of the ovary and the endometrium, can also be applied to breast cancer sections. They have to be evaluated on more patient material.

The image processing methods for breast cancer sections were based on the methods developed for tumors of the ovary and endometrium, described in detail in<sup>16-19</sup>.

In the 'Feulgen' method, it was assumed that the epithelial nuclei are more tightly packed and more darkly stained than the stromal nuclei. As this assumption does hold for breast cancers of the ductal type, the method equals the corresponding ones for ovarian tumors and endometrial carcinomas, except for the tuning of some parameters, as described in the 'Material and Methods' section. The resulting method for breast cancer sections consists of a mixture of parameter values used for ovarian tumors, such as the minimum pixel value in the threshold selection for the yellow image, and parameters used for endometrial carcinomas, such as the size of holes in the tissue and epithelium to be eliminated.

The 'anti-cytokeratin CAM 5.2' method for breast cancer sections equals the one applied to both ovarian and endometrial (pre)malignancies, except for the parameter describing the size of small objects and spaces after the segmentation of the epithelium. The size chosen has to be smaller for breast cancer sections, due to the presence of relatively many small epithelial and stromal segments.

It can be concluded that the greater part of the image processing methods for breast cancers, automatically estimating the epithelial percentage and selecting the most epithelium-rich areas, are equivalent to the methods for epithelial tumors of the ovary and endometrium. Therefore, having evaluated the algorithm on three different organs, it may be expected that these methods, after tuning the parameters of the tissue architecture to the specific characteristics of the new organ, apply to epithelial tumors of other organ systems as well.

#### Acknowledgements

The authors wish to thank Conny van Galen, Thea Tadema, Eric Noteboom, and Jos de Lange for their technical assistance.

#### References

- <sup>1</sup> Auer GU, Caspersson TO, Wallgren AS (1980) DNA content and survival in mammary carcinoma. *An Quant Cytol Histol* 2: 161-165
- <sup>2</sup> Baak JPA, Dop v H, Kurver PHJ, Hermans JO (1985) The value of morphometry to classic prognosticators in breast cancer. *Cancer* 56: 374-382
- <sup>3</sup> Dixon WJ (Ed.) (1983) *BMDP Biomedical Computer Programs*. University of California Press, Berkeley
- <sup>4</sup> Ewers SB, Langstrom E, Baltedop B, Killander B (1984) Flow cytometric DNA analysis in breast carcinoma and clinico pathological correlations. *Cytometry* 5: 408-419

- <sup>5</sup> Fisher ER, Sass R, Fisher B (1984) Pathologic findings of the national surgical adjuvant project for breast cancers. *Can* 53: 712-723
- <sup>6</sup> Fisher B, Redmond C, Foster R, Fisher ER (1985) Ten-year results of a randomized clinical trial comparing radical mastectomy and total mastectomy with or without radiation. *N Engl J Med* 312: 674-681
- <sup>7</sup> Haralick RM, Shapiro LG (1985) Image segmentation techniques. *Comput Vision Graph Im Proc* 29: 100-132
- <sup>8</sup> Kaplan EL, Meier P (1958) Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 53: 457-481
- <sup>9</sup> Langley FA, Baak JPA, Oort J (1983) Diagnosis making error sources. In: Baak JPA, Oort J (Eds.) *A Manual Morphometry in Diagnostic Pathology*. Springer Verlag, Berlin-Heidelberg-New York-Tokyo
- <sup>10</sup> Leader M, Patel J, Makin C, Henry K (1986) An analysis of the sensitivity and specificity of the cytokeratin marker CAM 5.2 for epithelial tumours. Results of a study of 203 sarcoma, 50 carcinomas and 28 malignant melanomas. *Histopathology* 10: 1315-1324
- <sup>11</sup> Linden vd JC, Baak JPA, Lindemann J, Hermans J, Meyer CJLM (1987) Prospective evaluation of the prognostic value of morphometry in primary breast cancer patients. *J Clin Pathol* 40: 302-306
- <sup>12</sup> Mantel N (1966) Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother Rep* 50: 163-170
- <sup>13</sup> Nemoto K, Vana J, Bedwani RN, Balur HW, McGregor FH, Murphy GP (1980) Management and survival of female breast cancer: results of a national survey by the American college of surgeons. *Cancer* 45: 2917-2924
- <sup>14</sup> Prewitt JMS, Mendelsohn ML (1966) The analysis of cell images. *Am New York Acad Sci* 128: 1035-1053
- <sup>15</sup> Sahoo PK, Soltani S, Wong AKC, Chen YC (1988) A survey of thresholding techniques. *Comp Vision Graph Im Proc* 41: 233-260
- <sup>16</sup> Schipper NW, Smeulders AWM, Baak JPA (1987) Quantification of epithelial volume by image processing applied to ovarian tumors. *Cytometry* 8: 345-352
- <sup>17</sup> Schipper NW, Smeulders AWM, Baak JPA (1989) Evaluation of automated estimation of epithelial volume and its prognostic value in ovarian tumors. *Lab Invest* 61: 228-234
- <sup>18</sup> Schipper NW, Smeulders AWM, Lange de JHM, Baak JPA (1989) Quantification of epithelial area by image processing applied to endometrial carcinomas. A comparison with ovarian tumors. *Hum Pathol* 20: 1125-1132
- <sup>19</sup> Schipper NW, Smeulders AWM, Lange de JHM, Baak JPA (1989) Automated estimation of epithelial volume by image processing, and evaluation of its prognostic value for ovarian and endometrial tumors (Subm. for publ.)
- <sup>20</sup> Schipper NW, Baak JPA, Smeulders AWM (1989) Automated selection of the most epithelium-rich areas in gynecologic tumor sections (Subm. for publ.)
- <sup>21</sup> Stenkvis B, Westman-Naeser S, Vegelius J, Holmquist J, Nordin B, Bengston E, Jarkrans T, Eriksson O (1979) Analysis of reproducibility of subjective grading systems for breast carcinoma. *J Clin Pathol* 32: 979-985
- <sup>22</sup> Stenkvis B, Bengston E, Eriksson O, Jarkrans T, Nordin B, Westman-Naeser S (1983) Histopathological systems of breast cancer classification. *J Clin Pathol* 36: 392-398
- <sup>23</sup> Weibel ER (Ed.) (1979) *Practical Methods for Biological Morphometry*. In: *Stereological Methods*, Vol. I. Academic Pr., London

Received December 4, 1989 · Accepted in revised form February 6, 1990

**Key words:** Image analysis – Epithelial percentage – Gynecologic tumors – Breast cancer

Nellie W. Schipper, Department of Pathology, Free University Hospital, De Boelelaan 1117, 1007 MB Amsterdam, The Netherlands