

Classification
Physics Abstracts
42.30.Sy

Automated Segmentation of Cytological and Histological Images for the Nuclear Quantification: an Adaptive Approach based on Mathematical Morphology

Abderrahim Elmoataz ⁽¹⁾, Philippe Belhomme ⁽¹⁾, Paulette Herlin ⁽²⁾, Sophie Schüpp ⁽¹⁾, Marinette Revenu ⁽¹⁾ and Daniel Bloyet ⁽¹⁾

⁽¹⁾ GREYC-UPRESA CNRS 6072, ISMRA, 6 bd Maréchal Juin, 14050 Caen Cedex, France

⁽²⁾ Laboratoire d'Anatomie Pathologique, Centre F. Baclesse, Route de Lion/mer, 14021 Caen Cedex, France

Résumé. — Une stratégie générale de segmentation combinant des critères locaux et globaux dans un processus de croissance dérivé de la ligne de partage des eaux est proposée. Les informations prises en compte sont de type contour-région ou couleur. La méthode est appliquée à différents cas d'images microscopiques de cytologie et d'histologie.

Abstract. — A general segmentation strategy allowing to blend multiple criteria such as contour-region or color information in a region growing process derived from the watershed transformation is proposed. It is applied onto different types of cytological and histological microscopic images.

1. Introduction

Image analysis offers to the pathologist a modern tool which can be applied to several problems in oncology : quantification of DNA content, quantification of immunostaining reactions, counting of nuclear mitosis, characterization of the tumor tissue architecture, *etc.* However, its introduction in clinical practice implies a complete automation and standardization of handling and goes together with the evaluation of the clinical interest of the parameters measured. One of the key steps is thus the segmentation of interesting objects. In most cases, it concerns the extraction of cellular nuclei with two main objectives: analysis of whole isolated nuclei (for DNA quantification), and *in situ* analysis of nuclei inside histological sections (for immunostaining quantification).

Several segmentation strategies can then be considered. In this paper, a general method combining contour and region knowledge is proposed and illustrated on two image families.

2. Materials

The two families used in this work come from acquisitions of microscopic images of breast and oesophageal cancers: monochromatic cytological images of whole isolated nuclei (Figs. 1a, 1b);

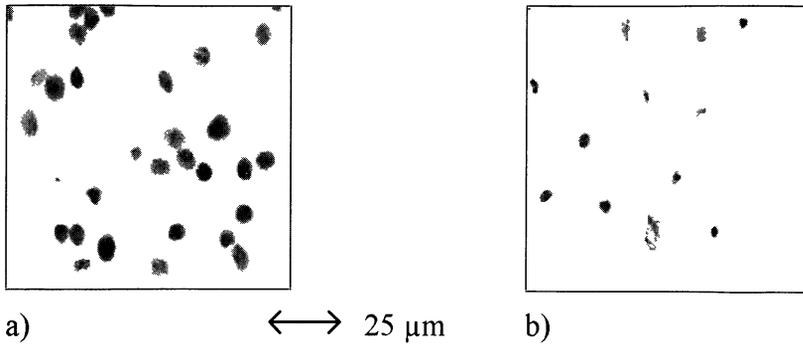


Fig. 1. — Examples of cytological images ($\times 125$): a) breast cancer; b) oesophagus cancer.

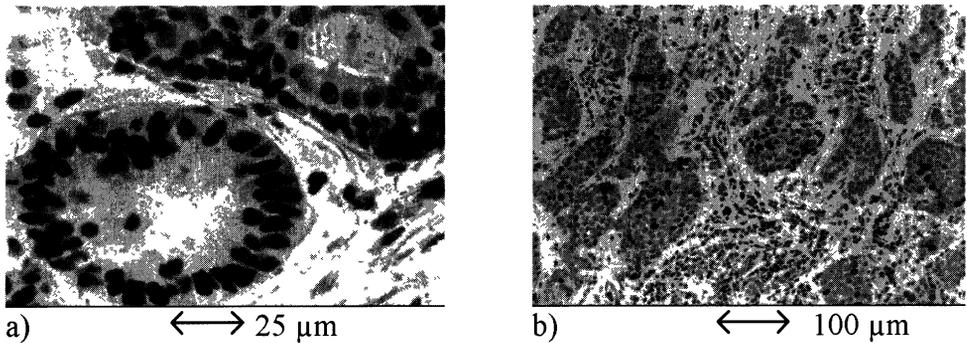


Fig. 2. — Examples of breast cancer histological images with two different magnifications: a) $\times 132$; b) $\times 33$.

color images of *in situ* nuclear profiles observed on histological sections of tumor tissues (Figs. 2a, 2b).

In the first family, nuclei are separated from their cytoplasm by an enzymatic digestion of thick sections ($50 \mu\text{m}$) of tumor biopsies fixed in formol and embedded in paraffin. Then they are placed onto slides and revealed by the DNA stoichiometric Feulgen staining. Images have a light background with relatively homogeneous texture, and dark objects (nuclei or debris) with mainly heterogeneous texture. Starting with this class, one wants to realize a reliable measure of nuclear DNA content which corresponds to the Integral Optical Density (IOD) computed from all the object areas. Because of the “glare effect”, the constraint is not to separate touching nuclei since the IOD should be mis-valued (even after a background correction and a controlled illumination of the slides, the glare-effect is still present). The segmentation procedure will have to retain the most objects as possible, and to provide a reliable localization of their borders in order to lead to an accurate measure of IOD and a reproducible evaluation of their intrinsic characteristics (shape, mean grey value, ...) for automatic classification.

In the second image family, acquisitions are run on $5 \mu\text{m}$ sections of immunohistochemically stained tissues. The analysis is aimed to quantify nuclear proteins (estrogen or progesterone receptors, proliferation markers) by revealing a staining bound to the associated antibody. A double diaminobenzidine-peroxydase and hematoxylin staining is processed onto sections after they have been placed onto glass slides. Diaminobenzidine and peroxydase are both combine

to reveal immunohistochemical markers and involve a brown coloration for positive nuclear locations. Hematoxylin is a DNA specific staining displaying a blue coloration for unmarked nuclei (negative locations).

Images of this class are more complex than in the previous case. One has to distinguish here many categories of objects: clusters of tumor cells (called lobules in carcinoma) and nuclear profiles presenting specific characteristics inside the clusters. So, two magnifications are used to complete this problem. A first one ($\times 132$) allows the study of intra-nuclear staining (Fig. 2a). A second one ($\times 33$) allows to observe tumor lobules (Fig. 2b). The goal of the analysis is to evaluate the immunostaining ratio defined as the positive nuclear area divided by the whole nuclear area, then to limit this measure inside the lobules. The automation of this procedure requires a precise segmentation at the tissue level (identification of lobules) as well as at the cellular level (identification of positive and negative profiles).

3. Methods

In order to define the best strategy adapted to these widely different problems, a segmentation frame divided in three steps is proposed:

a) A pre-processing step whose finality is to simplify images. It consists in regularizing images to soften some details having a small size. A simplification step can also lead to remove some objects according to either size criterion or contrast features.

b) A step of initialization of the segmentation process resulting in the extraction of object and background markers: this is an object-oriented step. Some image processing operators can take into consideration an *a priori* knowledge about the image family and about object features. They allow to extract the objects partially, so with an inaccurate localization of their borders. This step thus provides a splitting of pixels in two classes: labeled pixels belonging to objects and unlabelled pixels representing uncertain zones.

c) A step of localization that corresponds to the segmentation of the objects previously labeled. This step involves region growing processes using one or many criteria such as the watershed transformation [1], or its extension dealing with local and global information [2-4].

These three steps can be adapted to the problem to be resolved since they allow to consider various specific image processing operators and various segmentation strategies. Operators can be chosen in a base and their parameters tuned, then linked together in order to provide an optimal plan for the given task.

The scheme proposed here comes within the framework of the hierarchical morphological segmentation [5, 6] using operators from mathematical morphology which can take into account features such as shape, size, connexity or contrast.

4. Results

4.1 ANALYSIS OF CYTOLOGICAL IMAGES FOR DNA QUANTIFICATION. — The adopted plan is described as follow:

a) Image simplification by non linear diffusion [7]

$$u(x, y, 0) = I \quad (\text{initial image}), \quad u(x, y, t) = \text{Curv}(u) \cdot \|\nabla u\|$$

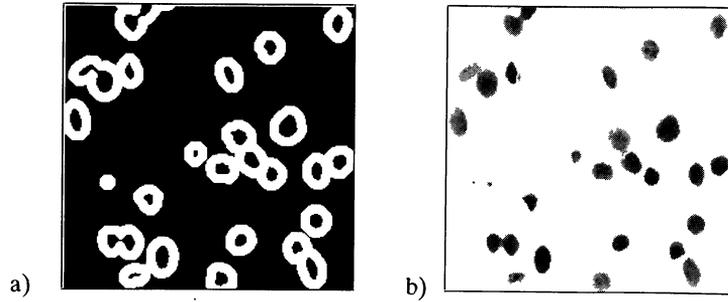


Fig. 3. — Segmentation of cytological images: a) labeled zones (in black); uncertain zones (in white); b) superimposed borders in white.

with

$$\text{Curv}(u) = (u_{xx} \cdot u_y^2 - 2 \cdot u_{xy} \cdot u_x \cdot u_y + u_{yy} \cdot u_x^2) / (u_x^2 + u_y^2)^{3/2}$$

$$\nabla u = (u_x, u_y) \text{ being the gradient of } u.$$

The simplification scale is set to $t = 0.6$ with a discretization step $\Delta t = 0.2$. This non linear smoothing operation has the advantage, over a linear smoothing, to lower noise effect while preserving object borders.

b) Extraction of markers

The background and nuclei are easily discriminated from their grey level distribution, so a simple binary threshold is well adapted. One has considered here the maximization of the inter-class standard deviation that consists in finding the threshold S which maximizes:

$$V(S) = N_0 \cdot N_1 \cdot (M_0 - M_1)^2 / N^2$$

with: N_0, N_1 , the numbers of pixels of each class, $N = N_0 + N_1$; M_0, M_1 , the first moments of the two classes.

The binary background marker is eroded by moving its borders away from uncertain zones. The object markers are eroded and labeled. One obtains an image M of markers more or less centered in nuclei (Fig. 3a).

c) Localization step

This step is completed by using the watershed transformation constrained by markers. This is done by imposing some minima in a gradient image g , computed from I , at the location of markers in M .

Starting with an image h such as: $h(p) = \begin{cases} 255 & \text{if } p \in M \\ 0 & \text{if } p \notin M \end{cases}$, a reconstruction by geodesic erosion of h into $V = \text{Min}(g, h)$ is computed [8] providing a new image $\varphi^{(\text{rec})}(V, h)$ onto which the watershed transformation is applied (Fig. 3b).

4.2 ANALYSIS OF HISTOLOGICAL COLOR IMAGES. — Two magnifications are used: the first one, $\times 132$, allows the study of intra-nuclear staining. The second one, $\times 33$, is used to distinguish tumor lobules and then to extract nuclei in them.

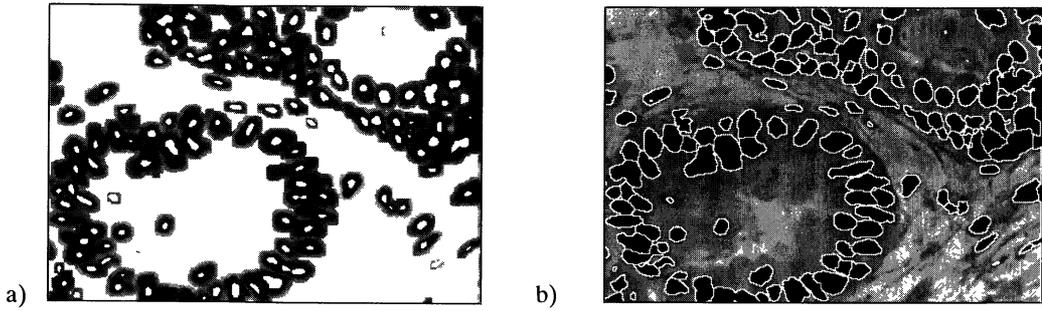


Fig. 4. — Segmentation of breast cancer histological images ($\times 132$): a) image of background and object markers; b) image of superimposed borders (in white).

4.2.1 *Magnification* $\times 132$. — The plan retained is:

a) Extraction of markers

The color image is split in its three RGB planes. The green plane I_G is used to extract both the background marker, by automated threshold, and the object markers, by computing $I_G - \varphi^{(\text{rec})}(I_G, I_G + h)$ (with $h = 10$) which provides the h -minima (Fig. 4a).

b) Localization of objects

A region growing process dealing with a local criterion (the gradient information) and a global criterion (the color information) blended according to the following formula:

$$C(p, \text{Re } g) = (1 - \alpha) \cdot \|M(I_{\text{Reg}}) - I(p)\| \cdot \|\nabla I(p)\|$$

with

$I(p) = [I_R(p), I_G(p), I_B(p)]$: vector of the color image I at the location p

$M(I_{\text{Reg}}) = [\bar{I}_{R,\text{Reg}}, \bar{I}_{G,\text{Reg}}, \bar{I}_{B,\text{Reg}}]$: mean vector computed for the region Reg

$\|\nabla I(p)\|$: norm of the color gradient image at the location of p

α : a weighting coefficient of the blending.

Figure 4b presents the result superimposed on the initial image.

4.2.2 *Magnification* $\times 33$. — To extract lobules and nuclei in these images, I_G which has the advantage of being a neutral component with respect to the two staining applied is also used (the color information only occurs for characterization of nuclei).

1) Segmentation of lobules

The tumor lobules are made of clusters which can be characterized by a small inter-cellular distance and whose nuclei have a greater size than the other cell categories (lymphocytes or stromal cells for example). The problem is that fibrovascular nuclei and lymphocytes nuclei are often gathered so that the distance criterion is not sufficient. Other features such as size or intensity are then to be considered. The plan retained is thus:

a) Image simplification in order to remove lymphocytes and to make clustering of other cells easier; this step makes use of morphological closing $\gamma_B(I_G) = \varepsilon_B \circ \delta_B(I_G)$, where δ_B and ε_B are the dilation and erosion by a plan structuring element B (of size 6 here).

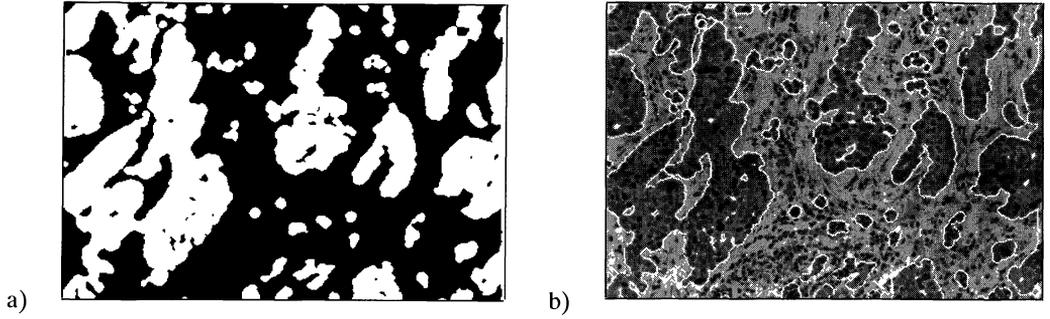


Fig. 5. — Segmentation of breast cancer histological images ($\times 33$): a) binary mask of lobules; b) borde of lobules superimposed onto the original image (in white).

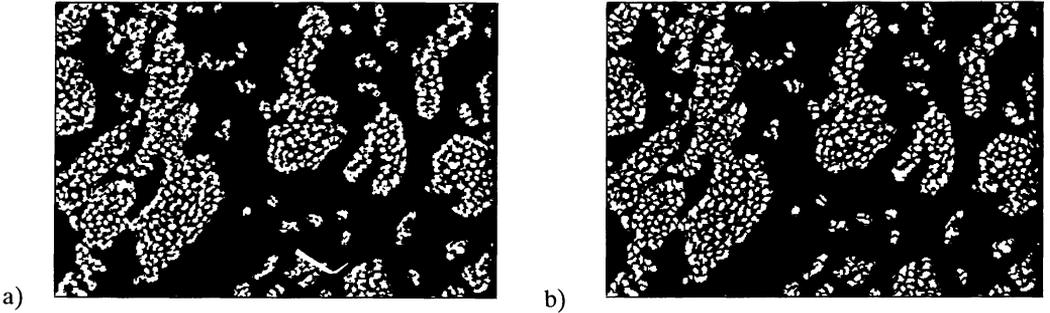


Fig. 6. — Segmentation of breast cancer histological images ($\times 33$): a) extraction of nuclei inside tumor lobules; b) splitting of touching nuclei.

b) Enhancing and automated threshold of the simplified image; indeed, the histogram of the enhanced image is bimodal; its two modes discriminate the background and lobule classes.

c) Closing of the binary image in order to remove smallest objects while connecting those that are close to clusters; the result is a binary mask I_L displaying the lobules (Fig. 5a) whose borders have been superimposed onto the original image (Fig. 5b).

2) Individualization of nuclei inside the lobules

a) Extraction of nuclei by residual analysis (providing a monochromatic image I_1 whose positive and negative values form a binary image I_2). In order to limit the process inside the sole lobules previously extracted, a logical intersection between I_2 and I_L is computed giving a new image I_3 (Fig. 6a). This step corresponds to the following operation:

$$I_1 = I_G - \text{Min}(I_G * G(x, \sigma), S) \quad \text{and} \quad I_3 = I_2 \cap I_L,$$

where $G(x, \sigma)$ is a Gaussian function of standard deviation σ , and S is the threshold estimated from I_G .

b) Splitting of touching nuclei

An inverted image of distance is computed from I_3 ; the watershed transformation is applied onto it (Fig. 6b).

5. Conclusion

A general segmentation method which we have applied to different cases of cytological and histological microscopic images has been presented. This method can of course be used with many other 2D images and, thanks to its principle, can also be extended to 3D images. Its capacity to solve some problems provided by confocal microscopy is currently evaluated.

Acknowledgements

This work has been undertaken under the auspices of the Pôle Traitement et Analyse d'images (TAI) de Basse-Normandie.

References

- [1] Beucher S. and Meyer F., *Mathematical Morphology in image processing*, chap. 12 (Marcel Dekker inc., New York, 1992) pp. 433-481.
- [2] Adams R. and Bischof L., *Seeded region growing*, *IEEE Trans. Pattern Anal. Machine Intell.* **16** (1994) 641-647.
- [3] Belhomme P., Elmoataz A., Herlin P. and Bloyet D., *Generalized region growing operator with optimal scanning: application to segmentation of breast cancer images*, to be published in *J. Microsc.* (1997).
- [4] Meyer F., *Color Image Segmentation*, IEE Fourth International Conference on Image Processing and its Application, Maastricht (1992) pp. 303-306
- [5] Coster M. and Chermant J.L., *Précis d'analyse d'images* (Les Presses du CNRS, Paris, 1989).
- [6] Serra J., *Image Analysis and Mathematical Morphology* (Academic Press, London, 1982).
- [7] Alvarez L., Lions P.L. and Morel J.M., *Image selective smoothing and edge detection by nonlinear diffusion (II)*, *SIAM J. Num. Analysis* **29** (1992) 845-866.
- [8] Vincent L., *Algorithmes morphologiques à base de files d'attente et de lacets : extension aux graphes*, Thèse de Doctorat de l'École Nationale Supérieure des Mines de Paris (1990).