



# Segmentation and classification of various types of cells in cytological images

Abderrahim Elmoataz, Marinette Revenu, Christine Porquet

## ► To cite this version:

Abderrahim Elmoataz, Marinette Revenu, Christine Porquet. Segmentation and classification of various types of cells in cytological images. International Conference on Image Processing and its Applications, 1992, Maastricht, Netherlands. pp.385 - 388. hal-00981836

**HAL Id: hal-00981836**

**<https://hal.science/hal-00981836>**

Submitted on 22 Apr 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## SEGMENTATION AND CLASSIFICATION OF VARIOUS TYPES OF CELLS IN CYTOLOGICAL IMAGES

A. ELMOATAZ, M. REVENU, C. PORQUET

LABORATOIRE D'INFORMATIQUE DE L'ISMRA, CAEN, FRANCE

INTRODUCTION

We are concerned with the segmentation of cytological images in order to automatically classify cellular categories identified thanks to their nucleus. We are dealing with microscopic images of human oesophagus cells, that have been dissociated and stained according to Feulgen and Rossenbeck methods (1). These images come from the pathology department of the cancer-research centre F. Baclesse of Caen. This study comes within the framework of the optimisation of an image analyzer dedicated to rapid detection of abnormalities of ploidy in human tumors (2) (3).

In this application, the purpose of segmentation is to extract cellular objects, to distinguish isolated cells from clusters and to calculate shape and texture measures to identify each kind of cells.

By using knowledge about these cytological images, we describe in this paper the results we have obtained in a preliminary study of segmentation and region classification into isolated cells and clusters.

First, we study the problem of segmentation of cytological images, and we lay the emphasis upon the knowledge involved in this kind of images and its exploitation in the selection process of segmentation operators.

Then we give a few segmentation methods based on thresholding, before describing in detail the segmentation operators we have selected in order to solve this problem.

SPECIFICATION OF THE PROBLEM

The images we are processing (fig. 1) are 256x256 images coded on 8 bits. They are characterized by cellular objects scattered on a background as well as clusters of cells and debris of cells. A cluster is composed of several cells touching each other, whereas debris correspond to isolated cells that have been truncated during the dissociation process.

Besides clusters and fragments, the following types of cells must be identified: stromal cells, epithelial cells and lymphocytes.

The images under study show specific shape and texture features:

- The background is characterized by a homogeneous texture and higher gray levels than the rest of the image.
- The cells have a convex circular shape, whereas the shape of a cluster of cells is non-convex.
- A normal non-textured cell shows a low inner contrast, whereas the inner contrast of clusters is high.

This a priori knowledge has led us to choose binary thresholding as an appropriate segmentation method in that case, and to select the three following measures for the classification of cells: the convexity degree, the elongation and the number of boundary pixels crossing a region.

IMAGE SEGMENTATIONImage segmentation: formal specification

Image segmentation is surely the most important stage in any image processing operation, but it is not the easiest to deal with. There is not indeed any sound theory or universal methodology. As a matter of fact, when one is faced with a given problem, on a specific class of images, with a precise segmentation objective, one chooses an adapted sequence of operators, the parameters of which are adjusted by the human expert in image processing. To select the appropriate sequence of operators, knowledge about the operators themselves must be used: parameters, activation conditions, results on certain images...

Many studies have been carried out on boundary detection, as well as on region-based segmentation (4) (5) (6) (7) (8) and many formulations have been brought up, but the best way of unifying these different approaches is to consider segmentation as a problem of finding a set of region and/or boundary primitives minimizing an energy function that takes into account various criteria on these primitives. This formulation leads us to a methodology to solve any segmentation problem following three independent steps:

- First step: Pre-segmentation. It consists in obtaining an initial segmentation by extracting region and/or boundary primitives adapted to the given problem. Various segmentation operators based on the boundary or the region approach can be used.

- Second step: Characterization. Intrinsic or relational features and measures on the region and/or boundary primitives extracted during the first step are calculated. They will be used to pass from one segmentation to another through the merging and splitting of primitives. That step must also involve evaluation or control measures of the initial segmentation.

- Third step: Search of the best solution by correcting the results of the initial segmentation and by using operators for merging or splitting region and/or boundary primitives.

In this paper, we are only describing some segmentation methods based on thresholding, because they are often used in the kind of problem we are interested in (9) (10).

## Image segmentation by binary thresholding

The idea is to consider that a region of an image is characterized by a gray-level interval ( $s_i, s_j$ ) and to detect all the pixels whose gray level is between these values. The region to be extracted thus constitutes a set of homogeneous gray levels producing a peak in the image histogram. To extract all the regions of the image, one has to consider histogram-analysis methods, in order to select the optimal thresholds  $s_i$  or to automatically detect the peaks of the histogram. Among these methods, one can mention those based on finding the maximum of the entropy function (11), on the preservation of the image moments, on discriminant analysis or on contrast. We are now giving some details about the two latter.

Thresholding by discriminant analysis. Let  $h$  be the gray-level histogram of an image of  $n$  pixels. The probability of a pixel to have value  $i$  is given by  $p_i = h(i)/n$ .

Let us suppose that the image is divided into two classes  $C_0$ ,  $C_1$  separated by the threshold  $s$ .  $C_0$  (resp.  $C_1$ ) corresponds to the pixels whose gray level is in  $[0, s]$  (resp. in  $[s+1, n]$ ). We define:

- The probability of pixel  $x$  to belong to  $C_0$  or  $C_1$ :

$$p(C_0) = \sum_{i=0}^{i=s} p_i \quad p(C_1) = \sum_{i=s+1}^{i=n} p_i$$

- The classes mean level:

$$M_0 = \sum_{i=0}^{i=s} p_i / p(C_0) \quad M_1 = \sum_{i=s+1}^{i=n} p_i / p(C_1)$$

- The total mean level of the original image:

$$M_t = \sum_{i=0}^{i=n} i \cdot p_i$$

Thresholding by discriminant analysis consists in maximizing the interclass variance which is classically defined by:

$$V(s) = \sum_{i=0}^{i=1} p(C_i) \cdot (M_i - M_t) \cdot (M_i - M_t)$$

Thresholding based on contrast. The problem is to determine thresholds ensuring the detection of the greatest possible number of high contrast boundaries and the smallest possible number of low contrast boundaries. Kohler (12) proposes to calculate contrast as follows:

Let  $x$  and  $y$  be two pixels next to each other of gray-level  $i$  and  $j$ . A boundary between  $x$  and  $y$  is detected by a threshold  $s$  if  $i \leq s < j$  or  $j \leq s < i$ .

The set of boundaries detected by  $s$  is:

$$K(s) = \{ \text{pairs of pixels } (x, y) / x, y \text{ neighbours} \\ \text{and } i \leq s < j \text{ or } j \leq s < i \}$$

The total contrast of boundaries detected by  $s$  is given by:

$$C(s) = \sum_{(x,y) \text{ of } K(s)} \text{abs}(s-i) + \text{abs}(s-j)$$

The mean contrast is:  $C_m(s) = C(s) / \text{card}(K(s))$ ,  $\text{card}$  being the cardinal of set  $K(s)$ .

The optimal threshold is the value maximizing  $C_m(s)$ .

## THE SEGMENTATION OPERATORS SELECTED

After this brief survey of some segmentation methods based on the region approach, in order to bring to the fore the specific knowledge about images used during the segmentation process, we detail the segmentation method we put forward. We proceed in three steps:

1. extraction of regions corresponding to isolated cells or clusters,
2. region labelling,
3. region classification.

### Region extraction

Two thresholding methods have been tested: the first one is a global method using boundary knowledge, whereas the second one is a local method using the Laplacian of the image.

Region extraction by binary thresholding. It consists in a direct delimitation of the cellular objects, based on the knowledge associated to the background which constitutes the complementary of cellular objects. This delimitation is achieved by an adaptative binary thresholding based on boundary knowledge.

The principle of our binary thresholding is similar to Kohler's method: one has to determine the threshold which detects the maximum of high contrast boundaries and the minimum of low contrast ones. But we use an original approach to automatically calculate that threshold.

This method can be summed up as follows:

- Image smoothing by using filters that preserve boundary information.
- Calculation of the Gradient image and estimation of the average amplitude of the Gradient image,  $AM$ .
- Automatic determination of the threshold giving a binary image, the boundaries of which best coincide with those of the Gradient image: For a given threshold  $s$ , we calculate  $NBP(s)$ , the number of boundary points, the amplitude of which is higher than  $AM$ . The optimal threshold is the one that maximizes  $NBP(s)$ .

Region extraction by using the Laplacian. The idea is to obtain closed boundaries from the zero-crossings of the Laplacian image. This method is very noise-sensitive and gives a lot of "false" boundaries as well as open boundaries. To make up for these drawbacks, we eliminate the noisy boundary points and try to close open boundaries.

The method can be summed up as follows:

- Calculation of an approximation of the Laplacian image and localization of all the zero-crossings.
- Elimination of the noisy boundary points.
- Closing boundaries by a conditional dilatation operator.

### Region labelling

The labelling of the binary image consists in searching the connected components. This is done by a simple scanning of the image. At the end of this operation, each region is associated to a single label.

After labelling the cells, we have to separate those which partially touch each other and those which are placed next to clusters. This is achieved in two steps:

- Enhancement of the labelled regions, either based on morphological operators or on methods using the boundary knowledge.
- Final labelling and dilatation of these regions.

### Region classification

During the classification stage, we scan the list of region labels and, for each region which does not touch the edges of the image, we calculate the following measures :

- the bounding rectangle of the region, in order to limit further operations to that only area.
- the surface.
- the convexity degree measured as the surface of the convex envelope of the region divided by the region surface. Getting the convex envelope of a region is done iteratively by the means of morphological operators.

During a first classification stage using only the convexity degree, regions are separated in two classes:

- epithelial cells and lymphocytes which are convex cells and thus have a high convexity degree,
- stromal cells, clusters and cells that have not been successfully segmented.

As stromal cells have an elongated shape, during a second classification stage, we can separate the latter from clusters by using the two following measures:

- the elongation,
- the number of boundary pixels crossing a region.

### RESULTS AND CONCLUSION

Figure 1 shows a cytological image containing various types of cells as well as clusters. Figure 2 shows the boundary image, obtained by an approximation of the Gradient, followed by a thinning operation and the elimination of lower amplitude boundaries. On figure 3 is shown the result of region segmentation by binary thresholding. On figure 4, we have only kept regions the convexity degree of which is higher than 0.9, the elongation higher than 0.7 and the number of inner boundary points lower than 10. These three values have been experimentally determined from cells that have been classified manually.

In this paper, we have presented a preliminary study on the segmentation of a class of cytological images and on a first classification of cells according to three measures. This work will be continued in four directions:

1. Improving the second segmentation method, by using operators to merge regions according to their average gray level.
2. Making a statistical study on segmentation errors in order to try to correct these errors by a re-segmentation of the regions classified as clusters based on a cooperation of other region extraction methods.
3. Trying to eliminate debris (truncated isolated cells) by finding a sequence of boundary points that constitute straight lines.
4. Improving the classification stage by adding texture measures.

### ACKNOWLEDGEMENTS

The authors are grateful to Mrs P. Herlin of the pathology department of the cancer-research centre F. Baclesse of Caen, who provided the cytological images and gave us all the explanations about the various types of cells we had to classify, and to Mr E. Masson for his help.

### REFERENCES

1. Feulgen R., Rossenbeck H., 1924, "Mikroskopisch-chemischer nachweis einer nukleinsäure vom typus der thymonukleinsäure und die darauf beruhende elektive färbung von zellkernen in mikroskopischen präparaten", Z. Physiol. Chemie, **135**, 203-249.
2. Masson E., Herlin P., Bloyet D., Duigou F., Mandard A.M., 1991, "Image analysis optimisation of the routine retrospective DNA ploidy measurement of solid tumors. Part I: Methodological choices", Acta Stereologica, in press.
3. Herlin P., Duigou F., Masson E., Bloyet D., Michels J.J., Mandard A.M., 1991, "Image analysis optimisation of the routine retrospective DNA ploidy measurement of solid tumors. Part II: Applications and output, Acta Stereologica, in press.
4. Chen M.H., Lee D., Pavlidis T., 1991, "Residual analysis for feature detection", IEEE Trans on PAMI, **13** (1), 30-40.
5. Munford D., Shah I., 1985, "Boundary detection by minimizing functionals", Proc. Computer Vision and Pattern Recogn, San Francisco, 22-26.
6. Gagalowicz A., 1989, "Coopération entre l'analyse et la synthèse d'images", 7ème congrès AFCET-RFIA, Paris, France.
7. Morel J.M., Soleimini S., 1989, "Segmentation d'images par méthodes variationnelles: une preuve constructive d'existence", Comptes rendus de l'Académie des Sciences, Paris, France.
8. Elmoataz A., Porquet C., Revenu M., 1991, "A general-purpose segmentation system using knowledge on images and segmentation operators", Advances in Intelligent Robotic Systems, SPIE vol. 1607, Boston, USA.
9. Coster M., Chermant J.L., 1985, Précis d'analyse d'images", Editions CNRS, Paris, France.
10. Chassery J.M., Garbay C., 1984, "An iterative segmentation method based on contextual color and shape criterion", IEEE PAMI, 360-372.
11. Zeboudj R., 1987, "Automated thresholding", 7th congress for stereology, Caen, France.
12. Kohler R., 1981, "A segmentation system based on thresholding", C.G.I.P., **15**, 319-338.

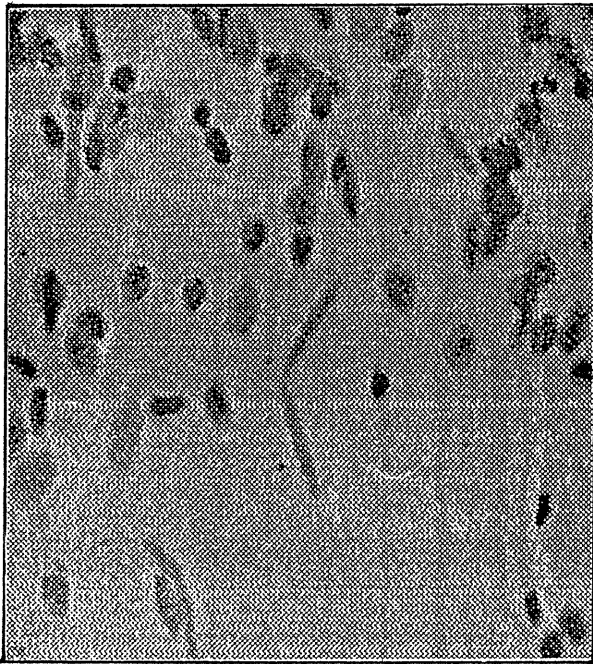


Figure 1: an image of cytocentrifuged deparaffined cells from normal oesophageal mucosae.



Figure 2: the boundary image of fig.1, obtained by an approximation of the Gradient, followed by a thinning operation and the elimination of lower amplitude boundaries.

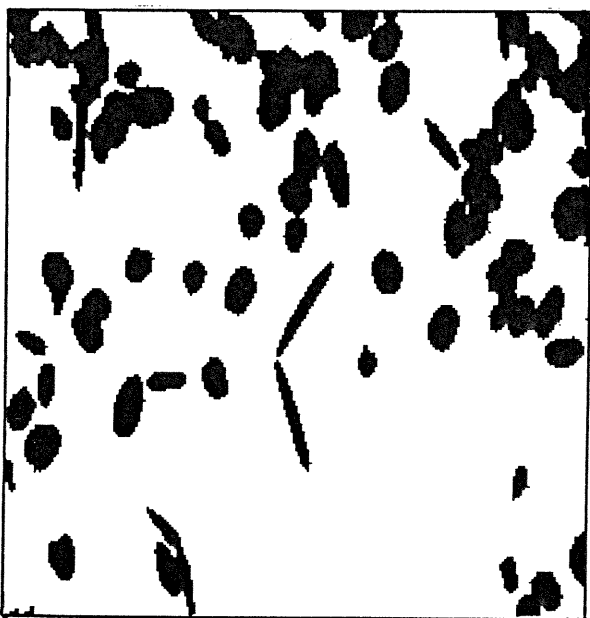


Figure 3: region segmentation by binary thresholding of fig. 1.

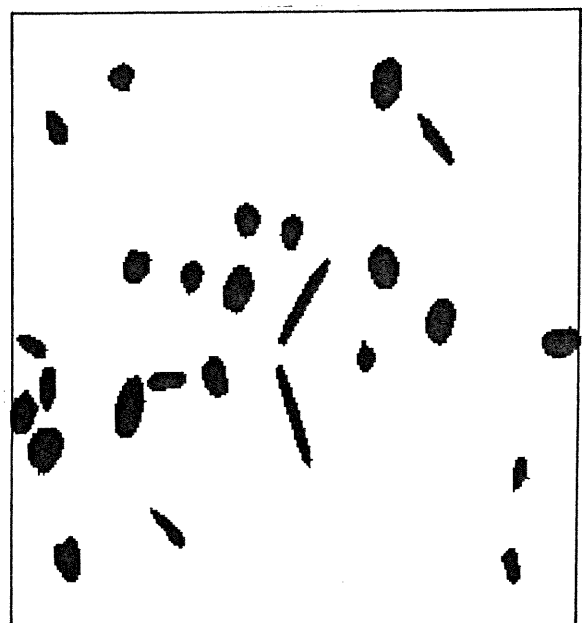


Figure 4: the regions of fig. 3 corresponding to isolated cells