

Morphological Description of Cervical Cell Images for the Pathological Recognition

N. Lassouaoui, L. Hamami, and N. Nouali

Abstract— The tracking allows to detect the tumor affections of cervical cancer, it is particularly complex and consuming time, because it consists in seeking some abnormal cells among a cluster of normal cells. In this paper, we present our proposed computer system for helping the doctors in tracking the cervical cancer.

Knowing that the diagnosis of the malignancy is based in the set of atypical morphological details of all cells, herein, we present an unsupervised genetic algorithm for the separation of cell components since the diagnosis is doing by analysis of the core and the cytoplasm. We give also the various algorithms used for computing the morphological characteristics of cells (Ratio core/cytoplasm, cellular deformity, ...) necessary for the recognition of illness.

Keywords—Cervical cell, morphological analysis, recognition, segmentation.

I. INTRODUCTION

THE interpretation of medical images is one of the most encouraging research fields, since it offers facilities for the diagnosis and the therapeutic decisions of a great number of diseases such as cancer.

The cervical cancer is a major problem of public health; it is the tumor affection most frequent at the woman. The only means of detection in time this illness, and consequently to decrease the death rate considerably, is the tracking. However, we realize that it is particularly complex and consuming time, because it consists in seeking some abnormal cells among a cluster of normal cells.

So, researchers thought of automating this work. This automatism consists in developing a system of assistance to the doctors by eliminating all the normal cases (by screening) and by presenting to the doctors only the abnormal cases by visualization on screen. Thanks to this system, we will be able to carry out an early detection of the illness, what will allow a more effective processing and will increase, consequently, chances of survival of the patient.

N. Lassouaoui is with Research Center on scientific and technical information, Mobile Computer Division, Rue des 3 frères aissou, Ben Aknoun, Algiers, Algeria (phone: 213-219-16204; fax: 213-219-12126; e-mail: nlassouaoui@mail.cerist.dz).

L. Hamami is with National Polytechnic School, Signal & Communications Laboratory, 10, Avenue Hassen Badi, El-Harrach, Algiers, Algeria, (e-mail: lhamami@mail.cerist.dz).

N. Nouali is with Research Center on scientific and technical information, Mobile Computer Division, Rue des 3 frères aissou, Ben Aknoun, Algiers, Algeria (e-mail: nnouali@mail.cerist.dz).

A computer vision system for tracking the uterine collar cancer is proposed [1], it will be a helped tool to the doctors by visualizing the cells and by given a decision about the malignity of the cells. It is composed by the various stages (Fig 1.), firstly, the biological cell image is obtained by a numerical camera of a microscope, this latter is connected to the computer. A pre-processing is applied to improve the quality of image and also to eliminate the useless information (such as: truncated cell) that can present a risk of mistakes in the following stages. After, we separate each cell to their cellular parts (core and cytoplasm) for analyzing their morphologies, knowing that the morphological characteristics of each cellular part give information about the malignity of each cell and the degree of illness. After, the system computes the morphological characteristics which are necessary in recognition stage for deducing a decision about the abnormality of the cells.

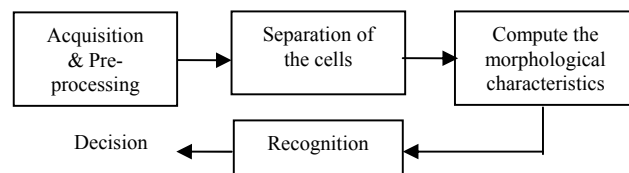


Fig. 1 Computer system for tracking the cervical cancer.

In this paper, we present our proposed algorithms for separating the cells and acquiring the knowledge necessary to the recognition of illness.

After presenting the characteristics of abnormal cells in following section, we present our developed algorithms for separating the cells and for computing the morphological characteristics, we give also an application result.

II. MORPHOLOGICAL CHARACTERISTICS OF ABNORMAL CELLS

The diagnosis of the malignancy is based in the set of the atypical morphological details of all cells, such as [2]:

- The most valid signs of the malignancy is the disruption of ratio core-cytoplasm in favour of the core, the cores of abnormal cells are usually thicker than those of the normal cells, but without that the cytoplasm presents a comparable variation.

- Nuclear deformity: Independently of variations of the size, there is a diversity marked in the shape of the cores of abnormal cells. They are more elliptic and queerly deformed.

- Cellular deformity: For an abnormal cell, the edges of its cytoplasm are deformed.

- We find also the no homogeneous textures of core and cytoplasm.

Since the malignity criteria appear on the morphology of the core and the cytoplasm of each cell, it is necessary to separate each cell on their components, that permits to analyze separately and easily their morphologies (size and shape).

In the following section, we present the algorithm used for separating the cell images.

III. SEPARATION OF CERVICAL CELL IMAGES

It is necessary in the separation of the cells to respect some characteristics, such as: The detection of the true edges of the core and the cytoplasm since their shapes and sizes give information about the illness. We precise that the cell image is divided mainly into three components: the core which is the heart of the cell, the cytoplasm which represents all the part of the cell that surrounds the core, and the background which surrounds the cells.

Several algorithms [3] are tested; however, we note that the segmentation of biological cell images is most difficult since the borders are often imperfect, stopped and supported by weak contrasts, badly definite. So, for a best detection of the cellular bodies, we segment the cervical cell images by an unsupervised selectionist relaxation algorithm. We have choose this genetic algorithm because it adapts to the spatial characteristics of pixels and detects even areas of intensity averages very close with those of their neighbors areas, so, we have a good separation even when two cells are near.

The principal aspect of selectionist relaxation algorithm [4] is the adaptation of unit to the position that occupies on image. It exists one unit $U=(B, L)$ for each pixel in image. The label L presents a symbolic manner of membership of unit to a given species. The chromosome B is a characteristic vector that constitutes a candidate representation of unit in environment. B is a matrix of luminous intensities.

There are two types of units: The *external* are in the periphery of units where at least one has a different label. The *internal* are in the center of units with the same label. The population of units evolves through a fine-grained distributed genetic algorithm (except the units of the limits of image); the units interact essentially with their 8 neighbors. We present briefly each step of algorithm.

Initialization: The labels of units are initially all distinct and they correspond to the position number of pixels in a sweeping line to line of image, so, each point is considered as a distinct area. Also, the parameters of characteristic vectors of units are uniformly generated in intensity interval of image (for an image of 256 gray levels, the parameters must be in [0, 255]).

Evaluation: The adequacy of each unit to the real characteristics of image in the neighbor is quantified by a fitness function. It measures the fidelity degree of each characteristic vector with the local characteristics of its

corresponding pixel in image. The fitness used is the Manhattan distance between the characteristic vector $B=(\beta_1, \dots, \beta_p)$ of unit and the vector $A=(\alpha_1, \dots, \alpha_p)$ determined at locality of the pixel s in image:

$$f_s(U) = -\sum_{i=1}^p |\alpha_i - \beta_i| \quad (1)$$

We precise that the label of each pixel is not used in the compute of f_s and the fitness must be stronger if the distance between the two vectors B and A is weak.

Selection: The tournament mode is selected and it is applied differently according to the label of unit:

- If the unit in a site s is internal, then select the unit of high fitness in neighbor of s .

- If the unit in a site s is external, then select randomly on image a unit U_r , then its fitness is compared to the fitness of all units in neighbor of s . If it is larger, then U_r is selected, if not select the unit of high fitness in neighbor of s .

Crossover: It exchanges portions of the chromosomes to obtain new ones. The generalised uniform crossover is applied with a probability p_c between unit in the site s and a random unit selected in its neighbor.

High value of p_c increases the recombination and the disruption of good strings [5]. Also, the search may stagnate for a low value of p_c . So, we take $p_c=0.5$.

Mutation: It insures diversity in genetic population. In [6], we have proposed to replace the allele by allele of gene randomly selected on the chromosome, this latter is selected randomly on the 8 neighbor chromosomes. Also, for decreasing the cumulated effect of mutation on N generations [6], the mutation probability Pm follows a decreased exponential law according to the generation.

We precise that external units are excluded from the crossover and mutation operators, since the units with different labels are supposed to be adapted to the distinct areas.

Stopping criteria: There exist several stopping criteria which can be the time, fitness, ... Here in, the number of generation is chosen. However, with inherent variability of the method and the contents of images, we do not have a means to find the optimal value of N , then, we propose to apply the algorithm for various values of N between [50, 250], and the most stable segmentation is retained.

By the fig 2, we give an application of the selectionist relaxation algorithm on cervical cell image. We note the ability of this method spatially adapted to detect areas with intensity averages very close with those of their neighbor areas (29 areas are detected for the fig 2.b). These areas are often not detected by several segmentation algorithms, because they are not sensitive to these weak variations of intensities.

However, the detected areas do not correspond to the cell objects necessary for recognition of illness. For that, we propose to classify by Fisher algorithm the detected areas by selectionist relaxation.

The Fisher algorithm determines the partition of N gray levels of image in k classes by minimizing the sum of the variances of the gray levels of each class [7]. So, we compute the two thresholds S_1 and S_2 which define the three classes of image, after, we classify each detected homogeneous area by selectionist relaxation to one class by comparing its intensity mean to S_1 and S_2 . Finally, we obtain the segmented image by affecting to each class an intensity, example, we affect the black intensity to areas classified to the class $[0, S_1]$.

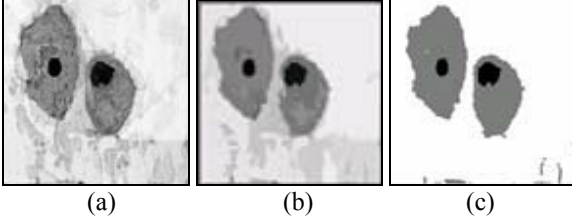


Fig. 2 Separation of cervical cell image (a), by selectionist relaxation (b) and by adding the fisher classification (c).

By adding the classification with the fisher algorithm (fig 2.c), the algorithm allows the detection the true classes (core, cytoplasm and the background) of cervical cell image, then, we can do the compute of morphological characteristics.

IV. COMPUTE THE MORPHOLOGICAL CHARACTERISTICS

A. Compute the ratio core-cytoplasm

The core of abnormal cell is thick; since we can measure this disturbance of the size by computing the ratio core-cytoplasm RCC which increases.

The epithelium includes four cellular layers, for each normal layer we have these RCC intervals:

- Basal cells: $0.25 \leq RCC < 0.37$
- Para-basal cells: $0.15 \leq RCC < 0.25$
- Intermediate cells: $0.04 \leq RCC < 0.15$
- Superficial cells: $0.01 \leq RCC < 0.04$.

If there is one type of cells in epithelium, we can easily compute the RCC by dividing the number of points in all cores on the number of points in all cytoplasm. Since, we have 4 types of the cells, we must find the RCC of each cell.

For that, we apply a second segmentation which affects a label to each cell area, so, the size of each cellular component is obtained by computing the number of pixels having its label.

In the literature [7], there are various labelling algorithms, in [8], we have presented an algorithm based on a resolution of equivalence. This method detects the vicinities between pixels and defines a label to the current point according to those of nearest points. There are two steps:

First Step: It consists in a sweeping of image line by line, affecting a temporary label to each pixel and computing the equivalent table which is a dynamic matrix.

So, for each point (i, j) of image Ima , with $(Ima[i,j]=1)$ do:
If the two predecessors $(i, j-1)$ and $(i-1, j)$ of the point (i, j) :

- 1) Are equal to zero, then we affect to $Ima[i,j]$ a new label $NLab$, we allocate a new column in an equivalent table to

$NLab$.

- 2) Have the same label Lab different to zero, then $Ima[i, j]$ is equal to Lab .
- 3) Have different labels $l1$ and $l2$ no equal to zero, then we affect the small label Lab_s of last labels in the columns of $l1$ and $l2$ in equivalent table. Lab_s is also inserted in the column of other predecessor label in the equivalent table. We precise that insertion is doing in a decreasing way, i.e., in the same column, the labels are classed from highest until the smallest.

Second Step: At the end of the first step, an area can have several labels, in this second step, the algorithm solves these equivalences by affecting to each label its equivalent label, this latter is the smallest one in the column of the label in the equivalent table T .

Finally, we do a second sweeping by affecting to each pixel in image its equivalent label.

By fig 3, we present a result of labelling of a cytoplasm image. The labelling algorithm separates the cytoplasm of each cell by affecting to each one an intensity (label), that permits to obtain the surface of each cytoplasm. The labelling algorithm is also applied on core image for computing the surface of each core, since we deduce the RCC of each cell (Fig 4).

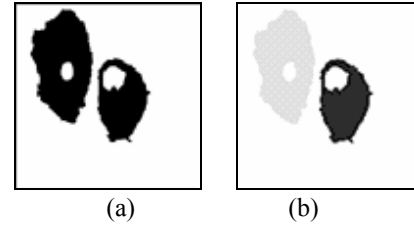
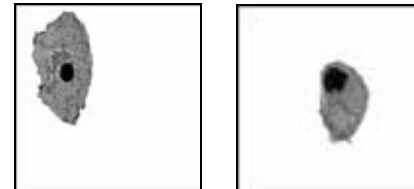


Fig. 3 Labelling of cytoplasm components.



(a) Intermediate $RCC=0.043$ (b) Para-basal $RCC=0.152$
Fig. 4 Classification by the compute of RCC of each cell.

B. Deformity of the cervical cell components

To measure the degree of deformity of each cellular component, we use the circularity factor f_c , with $0 \leq f_c \leq 1$. We precise that more f_c is near to 0, more the cellular component is deformed. Also, more f_c is near to 1, more the cellular component is circular. We note that the deformity in abnormal cell is more dangerous if it appears in the core.

The circularity factor is given by:

$$f_c = \frac{P^2}{S} \quad (2)$$

With: P is the perimeter and S is the surface.

For computing the circularity factor of the core and the cytoplasm, we detect the edges of each one by following the

limits of cell components obtained by the unsupervised selectionist relaxation (Fig 5).

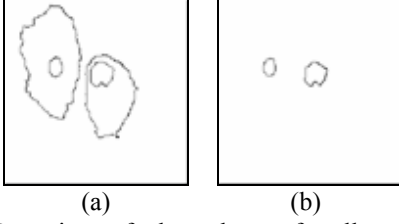


Fig. 5 Detection of the edges of cell components for computing the circularity factors.

C. Heterogeneous texture

For measure the heterogeneous texture of each component, we can use several and various parameters [9], here in, we present:

- The intensity mean: $MI = \sum_{i=1}^S I(i)/S$
- The variance: $\sigma^2 = \frac{\sum_{i=1}^S (I(i) - MI)^2}{S}$

With $I(i)$ is the intensity of pixel i and S is the number of pixels.

For measuring the heterogeneity in the core, we divide it on N small windows w_i of size 3×3 , and we compute the intensity mean $MI[w_i]$ and the variance $\sigma^2[w_i]$ of each window w_i . After, we compute the mean MI of intensity means and the mean of variances σ^2 of N windows. We compare the $MI[w_i]$ and the $\sigma^2[w_i]$ of each window w_i to respectively MI and σ^2 .

The window w_i is heterogeneous area if :

$$\begin{cases} |MI[w_i] - MI| > S_1 \\ |\sigma^2[w_i] - \sigma^2| > S_2 \end{cases} \quad (3)$$

With S_1 and S_2 are two thresholds.

We can also compute the anisocariose [2] which appears as the various pixels of intensities different to the texture of cell component. For detecting these points, we must choose an algorithm that works at level of the pixel, because if it works with the neighbor, the anisocariose pixels will be attenuated because the smoothing.

So, we chose the multifractal algorithm since it measures the texture and it works at level of the point. Herein, we present briefly the compute of multifractal parameter; more details are given in [1].

If the singularity exponent α exists at a point (i,j) , it is given by [1] :

$$\alpha(i,j) = - \frac{\ln(I(i,j)) - \ln(I_s)}{\ln(L)} \quad (4)$$

with :

- $I(i,j)$ is the gray level of the point (i,j)
- I_s is the measure of image of size L , it is the sum of intensities of all points.

After the compute of the singularity exponent of each point by using (4), we search the minimal singularity α_{min} and

maximal singularity α_{max} of image. Then, we search the three sub-intervals I_c, I_{cyl}, I_{back} of $[\alpha_{min}, \alpha_{max}]$, with each one corresponds respectively to the: core, cytoplasm and background. We note that a singularity interval $[\alpha_1, \alpha_2]$ corresponds to pixels with intensities in $\left[I_s \left(\frac{1}{L} \right)^{\alpha_2}, I_s \left(\frac{1}{L} \right)^{\alpha_1} \right]$.

We find the anisocariose in a core (respectively cytoplasm) by dividing N_c on its surface, with N_c is the number of pixels in the core (respectively cytoplasm) with singularities different of values in I_c (respectively I_{cyl}).

V. CONCLUSION

In this paper, we have presented the proposed computer system for helping the doctors in tracking the cervical cancer; we have principally give the various algorithms for designing the stage of detection the cellular components and the stage of computing the abnormal signs of cells.

We have used an unsupervised segmentation algorithm based on genetic operators with Fisher classification for separating the cells. This segmentation simplifies the analysis of each cell component for computing the illness characteristics.

We have also give the various algorithms used for computing these characteristics: the ratio core-cytoplasm, the nuclear deformity, the cytoplasm deformity, the heterogeneous texture of each cell component, anisocariose, ... We note that the obtained characteristic vectors of each cell are the input of the recognition stage of computer system of tracking the cervical cancer.

In future works, we design the recognition stage for deducing a decision about illness.

REFERENCES

- [1] N. Lassouaoui, L. Hamami, and A. Zerguerras, "Segmentation and classification of biological cell images by a multifractal approach", *International Journal of Intelligent systems*, vol. 18, n°6, 2003, pp. 657-678.
- [2] C. Gompel, *Atlas de la cytologie*, Medical school of Paris, 1982.
- [3] N. Lassouaoui, *Various approaches for image segmentation: Application to biomedical images*, doctorate thesis, Algiers, Algeria, 2004.
- [4] P. Andrey, *Image segmentation by genetic algorithms*, doctorate thesis, Paris 7, France, 1997.
- [5] K. F. Man, K. S. Tang, and S. Kwong, "Genetic Algorithms: Concepts and Applications", *IEEE Transactions on Industrial Electronics*, vol. 43, n° 5, 1996, pp: 519-534.
- [6] N. Lassouaoui, and L. Hamami, "A new mutation operator for the segmentation by selectionist relaxation of real images", *AMSE international conference on modeling and simulation MS'2004*, France, Jul. 2004, pp: 9.5-9.8.
- [7] R. C. Gonzalez, *Digital Image Processing*, second edition, Addison Wesley, 1987.
- [8] N. Lassouaoui, L. Hamami, and M. Lounis, "Classification of the cervical cell images by labelling algorithm", *Arab conference on Information Technology*, Egypt, vol. 2, 2003, pp: 682-693.
- [9] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification", *IEEE Transactions on systems, man and cybernetics*, vol. 3, n°6, 1973, pp. 610-621.