Classification of Cancer Cells Based on Morphological Features From Segmented MultiSpectral Bio-Images

A.CHADDAD, C.TANOGUAST, A.DANDACHE AND A.BOURIDANE
Laboratory of Interface Sensors and Microelectronics
Paul Verlaine University of Metz
7 Rue Marconi, Technopôle 57070 Metz
FRANCE
ahmad.chaddad@polymtl.ca

Abstract— In this paper a new approach aiming to detect and classify colon cancer cells is presented. Our detection approach was derived from the "Snake" method but using a progressive division of the dimensions of the image to achieve faster segmentation. Classification of different cell types was based on nine morphological parameters and on probabilistic neural network. Three types of cells were used to assess the efficiency of our segmentation and classifications models, including Benign Hyperplasia (BH), Intraepithelial Neoplasia (IN) that is a precursor state for cancer, and Carcinoma (Ca) that corresponds to abnormal tissue proliferation (cancer). Results showed that segmentation of microscopic images using this technique was of higher efficiency than the conventional snake method. The time consumed during segmentation was decreased to more than 50%. The efficiency of this method resides in its ability to segment Ca type cells that was difficult through other segmentation procedures. In classification only three morphologic parameters (area, Xor convex and solidity) were found to be effective to discriminate between the three types of cells. The results obtained using several images show the efficacy of the method.

Key-Words: - Classification, cancer cells, segmentation, multispectral bio-images.

1 Introduction
Multispectral image processing techniques are bringing new insights to biomedical research and other application domains in recent years. Pathologists make a diagnostic decision by viewing a specimen and measuring various diagnostically important attributes of an isolated object such as size, shape, darkness, color and texture. This is a complex process. In recent years, computer-aided image processing and analysis systems have played a significant role in quantitative pathology. The automatic recognition and classification of biomedical objects can enhance work efficiency while identifying new inter-relationships among biological features. The difficulty with the interpretation of colon biopsies is increasing because of the general trend in clinical diagnosis practice towards using smaller, less invasive sampling, often generating very small amounts of tissue. Therefore, the diagnosis tends to rely on cytology features, and hence a greater emphasis on nuclear characteristics. Almost all of the contributions in this area have focused on morphologic (shape) features [1]-[2]-[3]. In this work, we propose to use three features extracted from image as data to represent shape information of different multispectral cellular objects at multiple resolution levels. Then, the features are used in a simple rule-based classifier to discriminate between the three types of cancer cells [4]-[5]. Since each feature has clear physical meaning, the decision rule of this classifier is simple, which makes it very suitable for online processing. Research in medical images is accelerated when the manual search of abnormalities in medical images is replaced by an automatic search. Effectiveness of automatic search is analyzed and interpreted a large number of medical images in short time. This work is part of a project on the segmentation of colon cancer cells within multispectral image [5]. The objective of the present paper was to develop a new approach aiming to detect and classify colon cancer cells. Our approach was derived from the "Snake" method but using a progressive division of the dimensions of the image to achieve faster segmentation. Classification of different cell types was based on several morphological parameters and on probabilistic neural network. Three types of cells were used to assess the efficiency of our segmentation and classification models, including Benign Hyperplasia (BH), Intraepithelial Neoplasia (IN) that is a precursor state for cancer, and Carcinoma (Ca) that
corresponds to abnormal tissue proliferation (cancer).

2 Active Contour Segmentation

Active contour is a dynamic curve which tries, in an iterative process, to move towards and detect the contour of an object observed in an image. This curve consists of a set of points connected to each other by lines. An energy function is generally associated with this curve (snake) in such a way:

\[ F_{\text{snake}} = F_{\text{internal}} + F_{\text{external}} \]  

(1)

where \( F_{\text{internal}} \) is an energy that depends on the physical properties of the contour (\( F_{\text{connectivity}}, \ F_{\text{curvature}}, \ F_{\text{balloon}} \)) and \( F_{\text{external}} \) is another energy that depends on the properties of the image (\( F_{\text{grad}}, \ F_{\text{intensity}} \)) [6]-[7]. The corresponding algorithm tries to find a combination between different image points to minimize the energy function \( F_{\text{snake}} \) in order to detect and compute the contour. In our problem we have chosen a developed model of active contour which is able to detect the contour of an image without calculating its gradient and also without detecting its edges [8]. Let us consider Figure 1 representing an image of a single object having an intensity which is different from that of the background.

![Image 1](image1.png)

Fig. 1: Image representing a single object with uniform intensity \( U_{\text{internal}} \), separated by the contour \( C_0 \) from its background with uniform intensity \( U_{\text{external}} \).

Let \( c_1 \) and \( c_2 \) be the average intensities in the regions within and outside \( C_0 \), respectively. Energy \( F \) is defined as:

\[
F = F_1(C) + F_2(C) = \sum_{\text{interior}} |U_0(x,y) - c_1|^2 + \sum_{\text{exterior}} |U_0(x,y) - c_2|^2.
\]

(2)

According to \( F \) we can have four different cases of detected contour as shown in figure 2.

![Image 2](image2.png)

Fig. 2: Four different cases of detected contour.

By taking the four different cases of figure 2, we combine all possible outcomes of the green contour with the real one [4]:

\[
\begin{align*}
(a) & \quad F_1(C) > 0 \text{ and } F_2(C) \approx 0 \\
(b) & \quad F_1(C) \approx 0 \text{ and } F_2(C) > 0 \\
(c) & \quad F_1(C) > 0 \text{ and } F_2(C) > 0 \\
(d) & \quad F_1(C) \approx 0 \text{ and } F_2(C) \approx 0
\end{align*}
\]

(3) (4) (5) (6)

We can also add to the expression of \( F \), a normalization term which leads to the final \( F \):

\[
F = \mu \cdot \text{length}(C) + \sum_{\text{interior}} |U_0(x,y) - c_1|^2 + \sum_{\text{exterior}} |U_0(x,y) - c_2|^2
\]

(7)

where \( \mu \geq 0 \) is a fixed parameter. Therefore, the problem of contour detection can be simplified to the following equation:

\[
\inf_{c_1, c_2, C} (c_1, c_2, C)
\]

(8)

The multi-spectral images are then segmented at medium wavelength between 500 nm and 650 nm. These multispectral bio-images are obtained from an high-throughput Liquid Crystal Tunable Filter (LCTF filter) allows the capturing of 16 spectral bands between 500 nm and 660 nm with a 10 nm stepwise between each successive band. Thus, there are 16 sequential bands images for each type of malignant cancer grades (\( BH, IN \) and \( CA \)) for different frequencies or wavelength between 500 and 660 nm.

3 Adaptation of The Model for More Efficient and Faster Segmentation & Morphology Parameters

In this paper, we propose a new technique to adapt and use active contour model. This technique leads to an effective and faster segmentation of medical images especially those of high dimensions and resolution.

3.1 Eight Stages of the Segmentation Process

The first part is a set of eight consecutive stages of segmentation of the texture image. In the first stage,
the dimensions of the image are reduced to a value of 12.5% of their initial values. This means that for an initial image of 512 x 512 pixels of size, the image of the first stage I₁ is obtained by reducing the initial image to a size of 64 x 64 pixels (Fig. 3(a)). Given that the initial contour is a square of dimension 42 x 42 pixels, the active contour segmentation described above starts. Satisfying results are obtained after 2 to 4 seconds. This means that the algorithm could detect the true contour of the reduced image I₁ in order of seconds. Let C₁ be the resulting contour of the first stage segmentation (Fig. 3(b)). In the second stage of the segmentation, the initial image is reduced to a size of 25% of its original size. So an initial image of 512 x 512 pixels is reduced to image I₂ of size 128 x 128 pixels and the contour C₁ (64 x 64 pixels) result of the first segmentation stage is resiz ed to 128 x 128 pixels and it is used as an initial contour to segment the image I₂. C₂ is the contour obtained after this segmentation and it is resized [8] and used to segment another reduced image I₃. We continue with the six remaining stages in the same operation to finally get the final segmented (Fig. 3(c)). At each stage, the initial image is reduced to the corresponding dimension (image I₂) and the contour obtained by the previous segmentation stage (Cᵢ₋₁) is resized to Iᵢ’ size and used as initial contour to segment it. Table 1 shows the segmentation algorithm procedure.

This progressive segmentation of the image is performed automatically. A condition can be imposed on the number of iterations necessary to reach convergence. In this case, a set of 20 consecutive iterations were allowed for each segmentation step. We measure for each step the 20 iterations step until:

\[ D_{s1} < \alpha D_{s1} \]

(9)

where \( \alpha \) is a parameter that depends on the sensitivity of the required segmentation and is specified by the user. However, an analysis of the curve obtained at each step can give an idea about the value of the parameter \( \alpha \) such as:

\[ \alpha = \min \left( \frac{D_{s0}}{D_{s1}} \right) \]

(10)

Figure 4 illustrates the results of segmentation of these cells using the classical snake method. This segmentation required more than 6 minutes for each image beginning from a 492 x 492 square initial contour. It is evident that the contour converges toward local minima which are far from the real contour of each cell.

The results of our new segmentation technique for the three images are shown in figure 3. Successful segmentations were obtained within less than 3 minutes for each image. These images show the
efficiency of our model in detecting active contours of irregular objects such as cancerous IN and Ca cell types.

3.2 Nine Morphology Parameters

Nine morphologic and characteristic parameters were used to classify segmented cells. These parameters include area and perimeter of the cell, Xor cell-circle, Xor cell-convex, Xor cell-rectangle, standard deviation of the positions of the contour points, deviation sum, eccentricity, and solidity of the detected cell.

Xor cell-circle operator is applied between the cell and a circle having the same area and center of mass as of the cell. Figure 5 shows the three steps required to determine this parameter. The area of the white region in figure 5 (c) illustrates the numerical value of the Xor cell-circle parameter.

\[
\text{Area} = \sum_{i=1}^{N} (X_i - \bar{X})^2
\]  

where N is the number of the contour points, Xi is the distance between a contour point i and the center of the cell and \(\bar{X}\) is the mean value of Xi. The distance from each point of the contour to the mean contour is summed to determine the deviation sum as follows:

\[
\text{Deviation Sum} = \sum_{i=1}^{N} |X_i - \bar{X}|
\]

To measure the eccentricity, we interpose onto the cell an ellipse to cover it. The eccentricity parameter is given then by the following equation:

\[
\text{Eccentricity} = \frac{\text{distance}(f_1, f_2)}{L}
\]

where \(f_1\) and \(f_2\) are the two foci of the ellipse and L is its major axis length. After making the segmentation of microscopic images of three types of abnormal cells (BH, IN and Ca), Each of these parameters [1]-[3] are measured directly on three segmented images of the three types of abnormal cells, only three morphologic parameters (area, Xor convex and solidity) were found to be effective to discriminate between the three types of cells, the three parameters are:

Area: The area of the BH cell is 123,924 pixels, whereas cells IN and CA respectively an area of 48,187 pixels and 166,866 pixels. The difference between the areas of three type cells is very interested for discriminated.

xorCell convex: For the computation of this parameter, we must first draw a convex envelop of the cell and then calculate the segmented area of the white region which is obtained by the operator "XOR" between the segmented cell and the convex. Figure 6 explains how to obtain this parameter is set for cell (BH) 7017.75 pixels for the cell (IN) 38,321 pixels, and finally to the cell (Ca) 58605.12 pixels.

Solidity: It is the ratio of pixels in the convex hull of the cell, which are also in the region of the cell. It is calculated as:

\[
\text{Solidity} = \frac{\text{Area (cell)}}{\text{Area (convex)}}
\]

The three cells have been used BH, IN and Ca was obtained respectively the following results: 0.947, 0.557 and 0.740.

3.2.1 Algorithm

The algorithm we developed includes two consecutive parts. It receives first the microscopic image of the biopsy, which is a sequential band image. Specification of the desired band of the image as well as the segmentation sensitivity allows the algorithm to perform the eight successive stages of segmentation, so to provide a final segmented image. Incomplete cells in the image are discarded by the algorithm. In the second part, the algorithm extracts the morphological parameters of the segmented cell that are used as input for the
probabilistic neural network to detect the type of cell. 18 images of biopsies of abnormal cells (BH, IN and Ca) were used. Images were segmented and observed on a midrange wavelength of 580 nm. The value of the parameter $\alpha$ was set at 5%.

### 4 Classification & Results

We have used a probabilistic neural network [9]-[10]-[11] as an effective method for classification using segmented cells in the image and the extracted three shape parameters (Area, cell Convex Xor, Solidity). In our work, the probabilistic neural network was trained with 45 images of known types as shown in figure 7 for use later to classify any new cell of a segmented image. The three shape parameters of the cell (area, cell-convex Xor, solidity) are therefore the input to the probabilistic neural network. Our classification method is applied to several different cell types; figure 8 (a, b, c) shows the position of three cells of unknown types and the results are very interesting.

![Fig.7: clusters of abnormal colon cells, BH, IN and Ca. this clusters are according to three parameters, Area, cell-convex Xor and solidity of segmented cells.](image)

The activation function is used to introduce non-linearity in the functioning of the neuron. The activation function of the probabilistic neural network is a function that measures the distance of unknown variable to all known class variables. The three cell types using these three parameters. Although the three types are well separated in three regions, some IN cells are close to BH cells because of shape similarity between the BH and IN.

In a ROC curve the true positive rate (Sensitivity) is plotted in function of the false positive rate (100-Specificity) for different cut-off points as shown in figure 8(d), Therefore the closer the ROC curve is to the upper left corner, the higher the overall accuracy of the classification between the three types of cancer cells.

![Fig.8: classify three cells of three types of abnormal cell a) BH cell, b) IN cell, c) Ca cell, d) receiver operating characteristic curve.](image)
5 Conclusion
This paper proposes a classification method of cancer cells based on shape features obtained from multispectral images. The computational analysis just requires three morphological coefficients based on active contour (snake) segmentation and allows a successful classification between the types of cancer. We demonstrate a possible automatic detection of cancer cells of Benign Hyperplasia Intraepithelial Neoplasia and Carcinoma by using the three features. Finally, our approach is very simple, exhibits attractive results and the method is useful in the morphology features between different histopathological images thus allowing for a fast cancer cells classification of microscopic bio-images. However, this method requires accurate images at the training stage in order to decrease the error in classification. However, the technique is fast. As future work, we propose to research for new features about pixel value measurements also we propose to deal with a hardware implementation to accelerate the computation features by deploying FPGA and for decrease the time of computation technology to achieve high performance in fast computation in real time processing.

References


