Morphological Analysis of Mammary Biopsy Images

Elisa Martínez Marroquín, Eugenia Santamaría, Xavier Jové, J.C. Socoró
Department of Communications and Signal Theory
Escola de Telecomunicacions La Salle. Universitat Ramon Llull.
P° Bonanova, 8. Barcelona 08022, Spain
Phone 2902426 Fax 2902416 Email ee03292@els.url.es

Abstract - In this paper we present an automatic system to analyse cells' nucleus in a given biopsy of mammary tissue which is cancerous. Identification and characterization of cells' nucleus provides enough information to diagnose a high or low grade cancer. This is a high interest differentiation for the pathologist to apply the correct therapy. Performing these measures by human observation is a hard, imprecise and subjective task. The algorithm we present processes the image in order to identify cells' nucleus over the rest of the tissue. Images are enhanced and segmented using morphological transformations. An ultimate erosion is used in two steps to separate cells' nucleus in contact. It is based in a combination of symmetrical ultimate erosion with directional ultimate erosion.

I. INTRODUCTION.

In this work we present an automatic system to analyse mammary cancer images. High resolution optical microscopy provides the images from a transversal biopsy of mammary tissue. There are two principal groups of pathology: high damage level and low damage level. But there are biopsies which have some features of each group and deciding the group to which they belong is not trivial. The aim of the analysis is to supply parameters for a further fuzzy classification.

Nowadays the classification is done through human observation. When someone analyses a biopsy, his conclusions are based on a general perception of the image and are affected by several factors like experience, fatigue and even subjective feelings. Thus the percentage of agreement among qualified personnel is not always satisfying. An objective and more accurate method has become necessary. Image analysis is able to calculate measures, such as irregularity form factor or grey level variance in each cell, that a person would not be able to calculate. We propose these measures to be used as objective parameters to achieve an efficient classification. On the other hand, an automatic system relies human from the tedious work of studying each biopsy.

The goal of the analysis presented here is to identify each cell and afterwards measure its shape, size and grey level features. The evaluation of this features for each cell makes it possible to elaborate an statistic of the whole image. These data will make a further automatic classification possible.

The object of the processing is to isolate each cell. There are several clusters of cells' nucleus in contact or even overlapping each other. We use mathematical morphology transformations [1][2][3] to classify biopsy scenes according to cells' nucleus shape, size and grey level features.

The analysis is organized into four steps: image enhancement, segmentation, features extraction and fuzzy classification. In this paper, we present only the first two points. We divide the segmentation process in three steps: simplification, markers extraction and decision.

Simplification is treated in section 3. The marker extraction step is described in section 4 and decision is introduced in section 5. Finally, conclusions are drawn in section 6.

II. IMAGE FEATURES.

Unfortunately, the image is of rather poor quality due to the nature of the biopsy. The imperfections in the acquisition method are difficulties added to the noise implicit in biomedical images. We divide the image in three components: noise, background and foreground.

We consider noise every error source due to the image digitalization, as well as imperfections in the biopsy such as bubbles of air between the tissue and the glass in which it is put. We identify these bubbles as small circular or elliptical dots of high grey level. Segregations are also noise that introduce oscilations in the grey level in the space between the cells' nucleus. There can also be empty spaces whitin the image. Bubbles and segregations have no interest in the analysis, therefore they are treated as noise and have to be removed to achieve the isolation of cells' nucleus.

There is one more difficulty in these images; it is necessary to dye the biopsy in order to visualize the cells' nucleus. The dye varies from one image to another; what's more, it varies depending on the region of the image. Thus it makes no sense to work with absolut grey level criteria.

Cells' nucleus generally have a circular or elliptical form, but since we treat with anomalous cells' nucleus they can present irregular forms. The form is a parameter that will help us to classify the biopsy. The size and the degree of variation in the grey level are also of great interest. Tissues that belong to the lowly damaged group have cells' nucleus of similar sizes, uniform grey level and circular or slightly elliptical form. On the other side the ones concerning to the highly damaged group have cells' nucleus of different sizes, some of which are bigger than the normal ones, have a wide range of grey levels and may present irregular forms. Density of cells' nucleus varies not only between different images but also in the same image; there can be isolated cells' nucleus as well as cells' nucleus in contact or overlapping ones. It is necessary to find the number of cells' nucleus in these clusters and evaluate its form, size and grey level independently. The density of cells' nucleus in the image is also a discriminant parameter.

III. PREPROCESSING.

A. Simplification

Pictures in the original form are not suitable for image

0-7803-31-09-5/96/$5.00 © 1996 IEEE

analysis. It is necessary to clean the image in order to remove noise components and smooth the wide range of grey levels not only in the background but also inside the cells' nucleus. First of all we simplify the image using an alternate sequencial filter (ASF). All ASF [2][3] are based on the composition of two primitive families. We have applied a succession of opening and closing both with reconstruction. The applied filter is:

$$(open \cdot close)^{5}_{rec} = M_{n}^{rec}(M_{n-1}^{rec}(\dots M_{1}^{rec}(I))) \qquad (1)$$

where I is the original image and

$$\begin{aligned} M_{i}^{rec}(X) &= open_{i}^{rec}(close_{i}^{rec}(X)) \\ open_{i}^{rec}(X) &= dil^{\infty}(ero(X,i),X) \\ close_{i}^{rec}(X) &= ero^{\infty}(dil(X,i),X) \end{aligned} \qquad (2)$$

with ero(X,i), dil(X,i) being the erosion and dilation of X with an structuring element size i respectively, and $dil^{\infty}(Y,X)$, $ero^{\infty}(Y,X)$ being the dilation and erosion of Y by reconstruction of X. For more information about these morphological operators refer to [1].

The reconstruction [2] is necessary in order to keep the contours. The first iteration starts with an unitary size structuring element. The structuring element size increases every iteration. We stop the process when the structuring element size is that of the smallest cell we want to preserve. The result is an image where the noise has been reduced and the cell's inside has been equalized.

The target is now to equalize the background. A morphological TopHat transformacion [1] is used to eliminate the nonuniform background. Substracting the opening from the original image leaves the desired cells' nucleus and some residues of low grey level and small size. These artefacts are eliminated applying an opening of size three. The global operation is:

$$\begin{aligned} &open_{3}(TopHat_{10}\ (I)) \\ &TopHat_{10}\ (I) = I - open_{10}(I) \end{aligned} \qquad (3)$$

and $open_{10}(X)$ is the morphological opening of image X with structuring element size 10.

The only parameter in the TopHat transformation is the size of the structuring element for the opening. It corresponds to the size of the largest cell. However, experimental results have shown that this parameter is not critical at all; excellent results are obtained even when we use an structuring element twice the size of the largest cell.

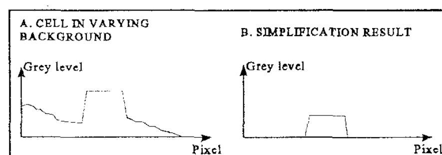

Fig. 1. 2D representation of an image line and its simplification result.

B. *Focus and selection of regions*

The useful scene is not the entire image plane. Firstly, there are different focus planes in the biopsy image, they lead to an image in which there are many unfocused cells' nucleus. These cells' nucleus are useless in our analysis because their form and grey level are not perfectly determined. We identify the unfocused cells' nucleus as those which present a weak and discontinuous contour in the morphological gradient image (Fig. 2). We eliminate these cells' nucleus and preserve the strong and closed contours. Then we enhance the regions enclosed by the remaining contours.

We threshold the gradient image in order to eliminate the weak contours. The threshold level is not critic at all. We fix this threshold experimentally observing several images. We obtain a binary image in which contours are black and the background is white. The binarized image presents the contours of the focused cells' nucleus as closed contours. On the other hand, contours belonging to unfocused cells' nucleus present discontinuities. In order to eliminate all discontinuous contours we look for all the white pixels connected with those in contact with the image frame. We expand the image frame pixels in a region growing way [4]. All pixels in this region are turned into black. The resulting image presents the interior of the focused cells' nucleus in a white colour over a black background. The result is a binary image with black background and white spots identifying the focused cells' nucleus regions.

Secondly, there are several cells' nucleus in contact with the image frame; these cells' nucleus can not be considered in the analysis due to its incomplete form and size. All objects touching the image frame are removed in order to avoid the analysis of incomplete cells' nucleus. We can see the resulting image in Fig. 3 which has been inverted to improve printing quality.

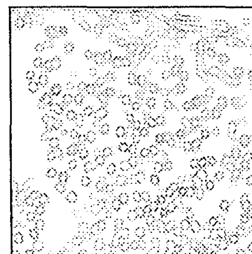

Fig. 2. Morphological gradient image.

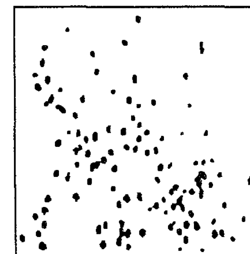

Fig. 3. Focus and selection of regions result image.

We reach a first approach to the location of the cells' nucleus but their form has been altered by the process. Moreover, without any further treatment it is not possible to distinguish between several overlapped cells' nucleus.

IV. MARKERS EXTRACTION

The aim of this process is to identify homogeneous regions and assign a marker to each one of them. Markers extraction is not concerned with contours location, which will be next section's goal. A marker is just a binary signal indicating the presence of an homogeneous area [1].

In grey images, frontiers correspond to significant changes in grey between two adjacent regions. These are detected by the gradient image, which is itself a grey tone function. If the original image is seen as a topological surface, then every object becomes a regional minimum in the gradient image.

This minimum is called a 'catchment basin', and it is surrounded by a so called 'chain of mountains'. It seems natural to consider the crest line of this mountains chain as the boundary and it is called 'dividing line'. Unfortunately, a problem subsists when applying this idea to real images. A cell has not a perfect constant grey value and thus will not imply one regional minimum but many of them, each one being surrounded by a small chain of mountains. The watershed algorithm [5] will detect all of them and will result in a severe oversegmentation. The solution to this problem is to select an smaller number of valid catchment basins. This means that we have to achieve only one catchment basin per cell; it is one connected area per cell. We call this area a cell marker. It is also necessary to define a marker to identify the background. The determination of the cells' nucleus markers forms a rough approach to cells' nucleus' location. It is indeed the intelligent part of the system. The next step is developed by the watershed algorithm which takes the marker and the gradient information to segment the image following a mechanical process.

A. Cells' nucleus markers. Division of joined cells' nucleus

In this section we attempt to find a procedure to isolate one marker for each cell. It is not possible to isolate the cells' nucleus from the background with a simple thresholding. It would be a hard task to decide the threshold level. And would be nearly impossible to avoid the problem of multiple markers per region.

The image obtained in section III marks the focused cells' nucleus regions but does not solve the problem with clusters of cells' nucleus that need to be divided. In order to solve this problem we apply an ultimate erosion [1][2] which will achieve a single marker for each isolated cell. Let us call this result E image. However, not all clusters are divided correctly; some of them still remain with a single mark.

Consequently, we will not extract all cells' nucleus at once, but start another phase in which we reconstruct every mark that can correspond to more than one cell. The process is sumed up in Fig. 4.

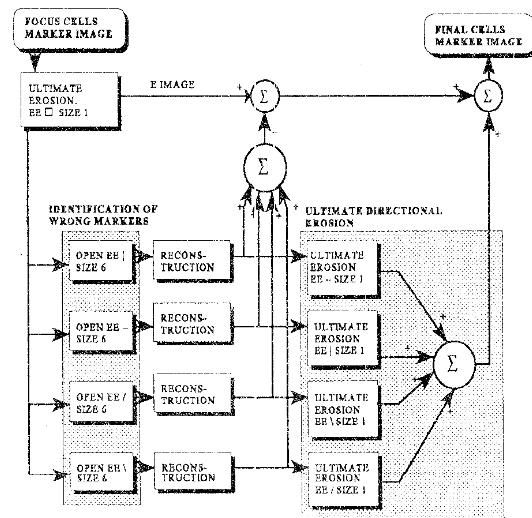

Fig. 4. Separation of cells' nucleus in contact.

Once we have these kind of markers reconstructed, they enter a process of ultimate directional erosion. Regions marked correctly present a single, small and circular mark, while the wrongly marked regions present a big non circular mark. We isolate this erroneous regions using a directional opening [2]. This is done by genereting a bank of four opening filters with a different oriented structuring element each: one per direction in the digital grid. These regions are reconstructed taking the E image as the reference. The reconstruction is substracted from the E image obtaining an image only with the markers properly found in the first ultimate erosion. Afterwards the reconstructed regions are reprocessed with an ultimate directional erosion that achieves the separation of the clusters.

We use the information of the reconstruction process to know the cluster's orientation. It determines the directional structuring element we will use in the second ultimate erosion. We use a longitudinal structuring element in the direction ortogonal to the direction of the structuring element corresponding to the filter that has detected the wrong marker. We add these new markers to the ones wich were properly found in the first ultimate erosion; the result is the cell marker.

B. Background marker

We extract the background marker from the image resulting of the focusing and selection of regions. This image has the cells' nucleus regions marked but does not identify the different cells' nucleus in a cluster. Nevertheless it is very useful to extract the background marker. It is necessary to achieve a single marker that represents all the background. We calculate the influence zones of each cell region and take the dividing line between zones as the background marker. The influence zone of a cell is defined as the set of pixels that are closer to that cell than to any other [1]. It is important to define the expansion of the zones with four connectivity in order to achieve an eight connected dividing line. The resulting dividing line is a single connected marker that identifies all the regions between cells' nucleus.

C. Grouping markers

The aim of the grouping markers process is to join all the information about homogenies regions. We add the cell markers to the background marker. It is important that no marker overlaps any other; they can not even touch each other. Should it happen the algorithm would take the markers in contact as a single one and the resulting image would show a subsegmentation. On the other hand it is also necessary that no marker contacts the contour of the cells' nucleus being analysed. If a marker touched a contour its expansion would incur in a wrong definition of the cell's fronteirs.

V. DECISION

In order to attain the ultimate goal of locating and classifying every particle, it is necessary to identify each cell independently and define its contorn accurately. This is achieved by applying the algorithm of watershed [5] to the gradient image starting the flooding by the marked regions. The application of the watershed algorithm is the mechanical

part of the system; however, the results depend on the images chosen as the inputs. Due to the noisy nature of the images, the determination of markers is fundamental to define the homotopy of the resulting image and then avoid an oversegmentation. Even using markers, if we apply the algorithm over the original image the result shows an erronious reconstruction of the cells' nucleus morphology. If it is applied over the simplified image, the reconstruction improves, but not enough. A satisfactory segmentation is achieved working with the morphological gradient image. The gradient enhance the fronteirs delimiting the expansion of the markers more accurately. We calculate the gradient over the simplified image resulting from the extraction of the background (Section III A). If we calculate the gradient over the original image the great noise component leads to a rough contour. On the other hand, if we calculate the gradient over an excessively simplified image, the resulting contour will not follow the correct form.

The decision step leads to an image where each cell is identified as an isolated element, the algorithm assigns a different uniform grey level to each cell and defines a single pixel width contour corresponding to the lines dividing regions.

VI. RESULTS AND CONCLUSIONS

Morphologic operators have proved to be very efficient for simplifying and segmenting. The proposed treatment proves to be efficient removing irrelevant components in the image and enhancing the interesting regions. It eliminates noise and background but does not alter or remove any cells' nucleus. Fig. 5 shows a biopsy image before processing; Fig. 6 presents the image resulting from the simplification step; Fig. 7 draws the segmented image and Fig. 8 corresponds to the edges extracted from the segmented image.

We have presented an algorithm to separate groups of cells' nucleus in contact. This algorithm is useful not only in clusters of cells' nucleus but also in any other application in which separation of any kind of particle is required.

The resulting image after the decision is formed by each cell over a uniform background. Each cell has an uniform grey level different from the level in the rest of cells' nucleus. Thus, identification of each cell separately is easily achieved. This makes it possible to measure cells' nucleus characteristic parameters such as area, perimeter, variation in grey level or irregularity of the form factor. All this parameters will make it possible to achieve a systematic classification based on objective factors.

The method gives an objective and systematic analysis that makes it possible to achieve the necessary parameters for a fuzzy classification.

REFERENCES

[1] E. Douherty, *An Introduction to Morphological Image Processing*, SPIE, 1992.

[2] P. Salembier and J. Serra, *Mathematical Morphology and its Applications to Image Coding*, UPC Course, 1995.

[3] R. M. Haralick, S. R. Stenberg and X. Zhuang, *Image Analysis using Mathematical Morphology*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 9 n° 4, July 1987, pp 532-550.

[4] W. K. Pratt, *Digital Image Processing*, John Willey & Sons, 1991

[5] Vincent. L. and Soille P., *Watersheds in Digital Spaces: An Efficient Algorithm Based on Immersion Simulations*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 13, June, 1991.

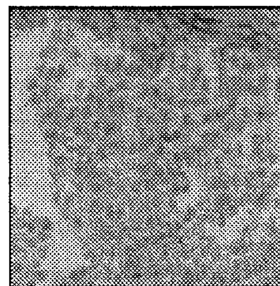

Fig. 5. Original image.

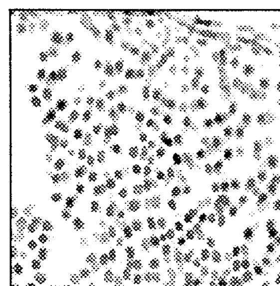

Fig. 6. Simplified image.

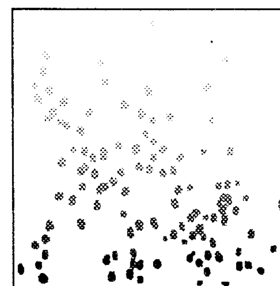

Fig. 7. Segmented image.

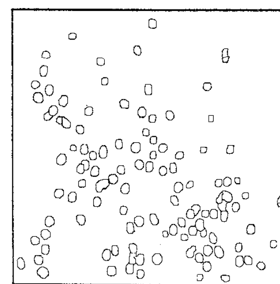

Fig. 8. Contour image.