

# **Detection of Ductus in Mammary Gland Tissue by Computer Vision**

Katarina Mele and Aleš Leonardis  
*Computer Vision Laboratory*  
*Faculty of Computer and Information Science*  
*University of Ljubljana*  
*Tržaška 25, 1000 Ljubljana, Slovenia*

## **Abstract**

In cell nucleus images of mammary gland tissue, both features of the nucleus and nuclei group shapes serve as cancer diagnosis criteria. Also, further treatment is dependent on the information about the cell nucleus spatial arrangement formed by malignant, potentially malignant, and normal ductus.

In this paper we present an automatic image analysis method that we developed for detecting the structures of nucleus clumps that represent boundaries of ductus. The method consist of the following stages applied in succession: segmentation with a threshold, greedy algorithm, relaxation, and graph search.

We tested our algorithm on images of mammary glands tissue. The results indicate that the method can distinguish between healthy ducts that have regular shapes and the malignant ducts (ductal carcinoma in situ) that have irregular shapes with boundaries appearing also inside the ducts.

This automatic procedure is a new approach in the area of cytometry and DCIS, which we believe will lead to a more reliable and objective evaluation of architectural characterization of nuclei group.

## 1 Introduction

While healthy cells are exactly programmed and form regular, similar and predictive shapes malignant cells grow out of control and spread to other parts of the body. We use the fact that shapes of normal structures are different than those that are malignant. We develop a method for the boundary detection of nucleus clumps and observe the difference between healthy and malignant clumps [1].

Morphological diagnosis is one of the most important diagnostic tool when considering patients with tumors [2]. The fact is that it is both cell nucleus features and nuclei group shapes that serve as cancer diagnosis criteria. So further treatment is dependent on the information about the cell nucleus spatial arrangement and the repeatability of the shapes formed by malignant, potentially malignant, and normal ductus. There are reports on the diagnosis and prognosis of breast cancer [3] and cervical cancer [4] with computer vision. All these systems are based solely on cell nucleus features.

Our inputs are images of cell smears that were stained by the Feulgen method [5]. The images were captured and digitized with an image cytometer. The first step in our method is a threshold segmentation to isolate the cell nucleus. The centers of gravity of isolated cell nucleus are used to form a fully connected graph. In the following steps we discard the graph connections that do not represent the boundary of the nucleus clumps. In the second step, different criteria are used to evaluate the connectivity and geometric properties. This step is based on local features while in the third step, the remaining links are assessed with an iterative relaxation algorithm. Iterative procedure propagates the influence of local features over a global area. In the fourth step we keep only those links that form long connected rounding paths. These paths represent the boundaries of the nucleus clumps. A brief sketch of the procedure is depicted in Figure 1.

We test our algorithm on tissue slice of breast with diagnosis Ductal Carcinoma in Situ (DCIS). This is intraductal carcinoma or non-invasive breast cancer DCIS which is confined to the ducts and has not invaded surrounding tissue. DCIS occurs when the cells that line the milk ducts in the breasts grow out of control but have not spread beyond the ducts.

Observing the results we notice that in the case of healthy ductus nuclei are distributed along the rim of the circular shape (Figure 10) in the case of malignant ductus they also appear inside the circular shape (Figure 8). Isolation of malignant ductus is a difficult task since the boundary contours are not apparent due to the cells spread inside the ductus.

The paper is organized as follows: The image analysis procedure is described in section 2. Section 3 includes some experiments. The conclusion is in section 4.

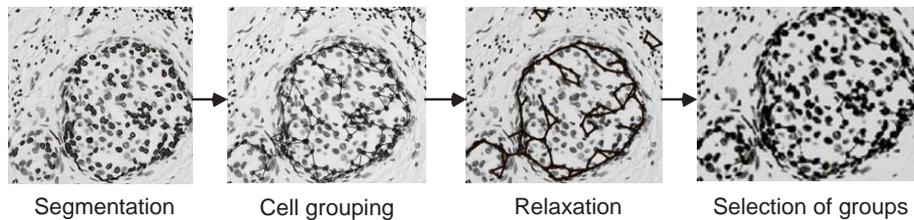


Figure 1: Outline of the method for the boundary detection of nucleus clumps.

## 2 Image Analysis Method

### 2.1 Cell Nuclei Segmentation

With segmentation we try to isolate nuclei that we subsequently use to form boundaries of the ductus. We use a simple thresholding method.

The problem with the simple thresholding procedure is that the method also detects the nuclei of lymphocyte. They also contain DNA and this is the reason why they are colored. All these particles represent noise and may disrupt further image processing. Nuclei of connective tissue represent noise as well. They appear in the neighborhood of the ductus. If we compare nuclei of ductus and nuclei of connective tissue we notice that nuclei of connective tissue are smaller. Thus, we eliminate a lot of that noise by an additional threshold as the size of segmented parts, which discards all segmented parts that are smaller than the specified threshold.

There is another problem with simple threshold cell nuclei segmentation. The cell nuclei are not uniformly colored so bright parts inside the nuclei appear that are treated as background after segmentation (Figure 2: left). These bright parts are either a consequence of diluted parts of chromatin (nucleus feature) or they appear during the preparation of nuclei for analysis. The problem is solved by merging these bright parts with nuclei (Figure 2: right).



Figure 2: Left: bright parts inside nuclei; right: bright parts are merged with nuclei.

## 2.2 Cell Nucleus Grouping

After the segmentation of individual nuclei we have to recover their possible spatial arrangements. Different criteria of perceptual grouping are used. The most useful criteria in our case are the distance between two nuclei and the criterion of angles. In this step we limit the number of all possible connections among nuclei.

### 2.2.1 Criterion of angles( $\mu_{ang}$ ):

The criterion of angles between nuclei is defined as:

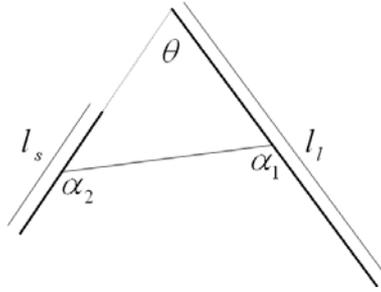


Figure 3: Criterion of angles

$$\mu_{ang} = \Sigma - \Theta \quad (1)$$

$$\Sigma = l_1 \cdot \alpha_1 + l_2 \cdot \alpha_2 \quad (2)$$

$$l_1 = \frac{l_s}{l_s + l_l} \quad (3)$$

$$l_2 = 1 - l_1 \quad (4)$$

$\Theta$  - angle between two segments,

$l_s$  - length of the shortest segment,

$l_l$  - length of the longest segment,

$\alpha_i$  - the largest angle of the two between the segment and the connection  $i=1,2$ ,

$\Sigma$  - sum of the largest angles  $\alpha_1$  and  $\alpha_2$  which are formed by a segment and the connection (Figure 3) and weighted by the lengths  $l_s$  and  $l_l$  of the corresponding segment.

Nucleus is represented by the center of gravity and the orientation. The criterion of angles gives more weight the connections (at the same angle  $\Theta$ ) that connect

more collinear objects. The criterion also considers the direction of the connection. The connections that go in the directions of largest (“more important”) object have high score. We support neighbouring nuclei in terms of orientation. This is important because several nuclei that are very close or even overlap are sometimes segmented as one. This nuclei usually lie in the direction of the boundary, actually they form the boundary. Individual nuclei on the boundary are in the most cases more flattened and tangential to the boundary. So we give higher score to the connections which are in the direction of the boundary and smaller to those that deviate from of that direction. The direction of the boundary in Figure 4 is marked with pattern line.

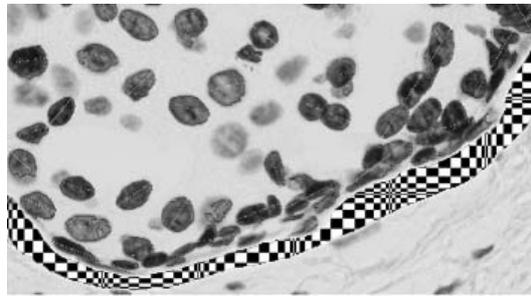


Figure 4: Boundary

### 2.2.2 Cell nucleus grouping procedure

Further cell nucleus grouping procedure has three steps.

1. First we connect all cell nuclei that are closer to each other than a threshold that was determined in advance.
2. Then we limit the number of connections to maximum 6 per nucleus. We estimate all the connections with the criterion of angles, sort them in a descending order, and add them into a new list using a greedy algorithm. We keep adding the connections into the list as long as it is possible to add a connection in a way that none of the nuclei has more than 6 connections.
3. At the end, the number of the connections is reduced again. The criterion function finds for each nucleus the two connections that form the largest obtuse angle. These two connections are assigned the weight  $w_1$ . We select the connections with the largest obtuse angle because they normally form the boundary we are trying to detect. If a nucleus has more than two connections the remaining connections are assigned the values from  $w_2$  to  $w_5$ . If a

nucleus has just one connection the connection has value  $w_2$ . Every connection connects two nucleus so it has two assigned weights — one from each nucleus. The overall weight of the nucleus is sum of these two. We again sort the connections in the descending order and according to the greedy algorithm select maximally 4 connections per each nuclei. The remaining connections form a graph that is an input for the relaxation stage.

### 2.3 Relaxation

Relaxation is an iterative procedure for global propagation of local connections [6]. It is defined by the rules with drive connection values converging either to the maximum (MAX) or to the minimum (MIN). The number of iterations (N) is fixed. We keep the connections that converge to MAX and discard the others. It is important that the relaxation spread the influence of a single connection over the whole graph although the connections in a single iteration step are valued locally.

The features used in the relaxation procedure are:

- the type of connection
- the rim lines and neighbors of the rim lines
- the border lines
- the connections with two strong neighbors

**The type of connection:** The type of connection  $e$  can be represented using a number pair  $i - j$  describing connection patterns at each vertex, where  $i$  and  $j$  are the vertex types of the connection  $e$ . Vertex type is the number of connections emanating from the vertex, not counting the edge  $e$ . Also the connections that are weak with respect to the other connections in the same vertex do not count. The vertex type can have values from 0 to 3. Three examples of the types of connections are depicted in Figure 5. The line thickness is proportional to the values of the connections.

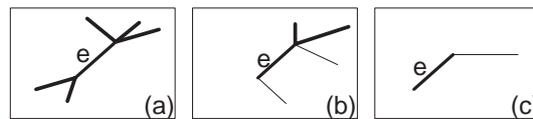


Figure 5: The type of connection (a) 2-3 (b) 1-2 (c) 0-1.

**Rim lines and neighbors of the rim lines:** If all neighbor connections of a connection are on the same side with respect to the straight line that goes through the connection then the connection is called a *rim line*. Figure 6 shows the rim line  $f$ . Thin line represent a straight line through the connection  $f$ . All other neighbor

connections are on the same side of the straight line. Connections  $f_1$  and  $f_2$  are the neighbors of the rim line because they form the largest angles  $\beta_1$  and  $\beta_2$  with the connection  $f$  on each side.

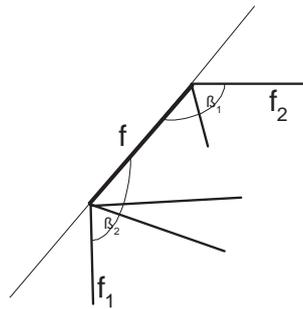


Figure 6: Rim lines

**Border lines:** On each side of the connection we draw a square. The side of the square equals the connection length. One side of the square is the connection itself. If at least one square is empty the connection is called *border line*. An empty square means that there is no connection or part of it in the square. An example of the border line  $g$  is depicted in Figure 7.

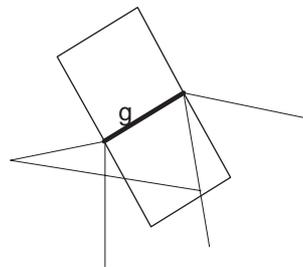


Figure 7: Border lines

**Connection with two strong neighbours:** Connection with two strong neighbours is a connection which has at least one neighbour on each side whose value approaches MAX. The angle between the connection and the strong neighbour must be at least  $\alpha$ .

### 2.3.1 Relaxation procedure

1. Calculate the initial values for all connections  $c^{(1)}(e)$ , using the criterion of angles.
2. Determine connections features.
3. Update the values  $c^{(k+1)}(e)$  of each connection according to its features and its previous value  $c^{(k)}(e)$ . The value can be increased (Eq. 5) or decreased (Eq. 6).

$$c^{(k+1)}(e) = \min(MAX, c^{(k)}(e) + \delta) \quad (5)$$

$$c^{(k+1)}(e) = \max(MIN, c^{(k)}(e) - \delta) \quad (6)$$

k - iteration number

$\delta$  - value for which we increase or decrease the value of connections

We increase the connection value if the connection is the rim line, the border line or has two strong neighbours. If the type of the connection is 0 – 0 or 0 – x (x is any value) we decrease the connection value.

4. Repeat the steps 2 and 3 until the number of iterations reaches the number N.

### 2.4 Selection of the envelope

Finally we select the most prominent envelopes generated at the relaxation step.

The aim of this step is to find the boundaries of ductus. One can see (Fig. 8 - 10) that the boundaries formed by cell nuclei are smooth and curved. These are precisely the structures we would like to select.

Curves are actually represented as paths in graphs. Paths are series of neighbour connections where angles formed by neighbour connection must be large enough.

The algorithm finds several very similar paths. This means they differ in a very small number of connections. They encompass almost the same vertices. We keep just the best according to an optimization procedure. We evaluate pairs of similar paths with function  $c_1$  and keep the one with the largest value of  $c_1$ .

$$c_1 = \frac{1}{n-1} \sum_{i=1}^{n-1} \Theta_i \quad (7)$$

n - the number of all connections in a path

$\theta$  - angle between two neighbour connections on the same path

v - value of a connection

### 3 Results

Figures 8-10 depict results for malignant, potential malignant, and normal ductus. We can see that our method can distinguished between healthy ducts that have regular shapes and the malignant ducts (ductal carcinoma in situ) that have irregular shapes with boundaries also appear inside the ducts.

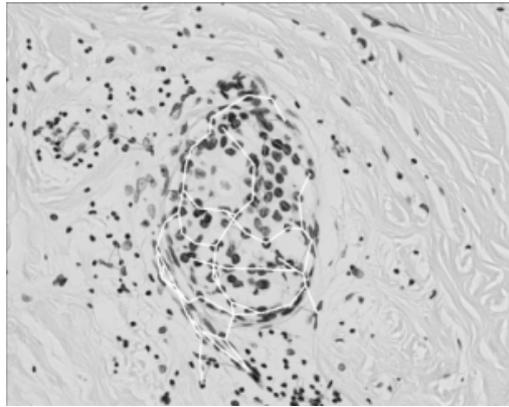


Figure 8: Malignant ductus

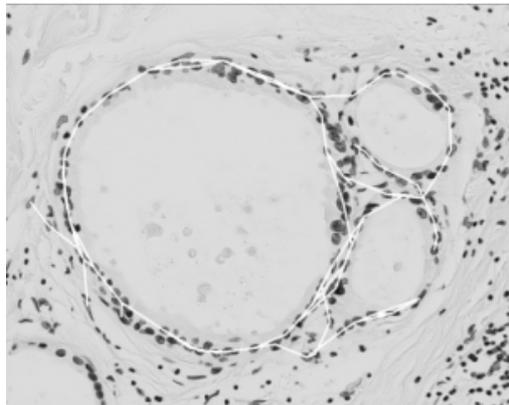


Figure 9: Potentially malignant ductus

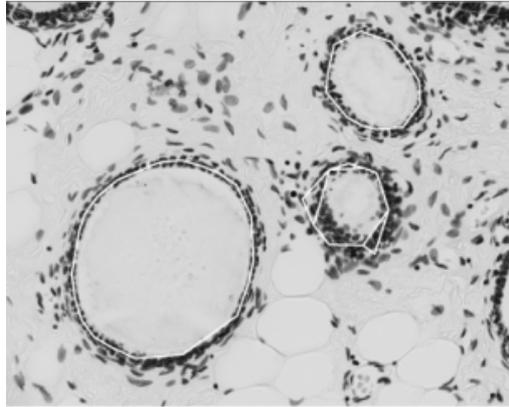


Figure 10: Normal ductus

## 4 Conclusion

Research that would combine architectural and nucleus features of DCIS is reasonable. It is a well-known fact that estimates of the architectural characteristics are hardly recurrent. Estimates of the architectural features are subjective. This automatic procedure is a new approach in the area of cytometry and DCIS, which we believe will lead to a more reliable and objective evaluation of architectural characterization of nuclei group.

## References

- [1] Mele, K., *Analiza slik celičnih jeder z računalniškim vidom*. Univerza v Ljubljani, Računalništvo in informatika: Ljubljana, 2000.
- [2] Lamovec, J., Patologija malignih tumorjev dojke. *Šola mamografske diagnostike*, **32(7)**, pp. 49–58, 1998.
- [3] W. H. Wolberg, W.N.S., Breast cytology diagnosis via digital image analysis. *Analytical and Quantitative Cytology and History*, **15(6)**, pp. 396–404, 1993.
- [4] Bamford, P. & Lovell, B., Unsupervised cell nucleus segmentation with active contours. *Signal Processing*, **71(2)**, pp. 203–213, 1998.
- [5] Lavrenčak, J., *Boundary Element Techniques*. Univerza v Ljubljani, Medicinska fakulteta: Ljubljana, pp. 18–25, 1998.
- [6] M. Sonka, V.H. & Boyle, R., *Digital Image Processing*. PWS, 1998.