Analysis of the Spatial Arrangement of Cells in the Proliferative Breast Lesions

Vincenzo Della Mea, Carlo Alberto Beltrami

Department of Pathology, University of Udine
33100 Udine - Italy

Abstract. A particularly difficult problem in the field of histopathology is the diagnosis among epitheliosis, atypical hyperplasia and carcinoma in situ.
This work is an attempt to reproduce the reasoning made by the pathologist when he/she analyzes the spatial arrangement of the cells in the duct, that is different depending on the grade of malignancy of the lesion. For this aim a method for the representation of the duct is needed, in order to identify the structures involved in the diagnosis.
A representation based on perceptual graphs is derived from an initial set of nodes that represents the cell nuclei of the lesion. On this representation the evaluation and quantification of structural features may be carried out.
A prototype was developed to evaluate the clinical significance of the approach. In a preliminary phase the statistical analysis of some features revealed significant differences among the three pathologies, encouraging the progression of the work.

1 Introduction

A difficult problem in the field of histopathology is the classification of proliferative breast lesions, i.e., the diagnosis among epitheliosis, atypical hyperplasia and carcinoma in situ. These three categories were first described by the College of American Pathologists [2] and by Page [9,11] and each of them was considered to have a different relative risk for the development of infiltrating breast cancer. In his work, Page presented a set of rules for the discrimination among the pathologies. These rules are divided in three different subsets, involving respectively cytologic features, architectural characteristics and anatomic extent of the lesion. The rules of the first subset are based on cytologic findings such as cytoplasmic aspect, nuclear shape, chromasia; these features have been investigated but the results are not satisfactory.

This paper focuses the attention on the rules belonging to the second subset, i.e., those capturing the way the cells occupy the space differently depending on the grade of malignancy of the lesion. The use of these rules implies a form of spatial reasoning on the lesion, with the aim of recognizing and quantifying spatial properties of the whole lesion or regions of it.
A modelization of the problem involving the use of perceptual graphs is proposed; more specifically, a hierarchy of graphs is used to represent the lesion.
The study of the structural characteristics of the lesions may offer some advantages: first of all, it guarantees the quantification of features hard to evaluate in an objective manner; then it offers the possibility to develop a system that avoids the classical problems related to morphometry and densitometry, and finally it gives the

opportunity to understand thoroughly the biological significance of proliferative lesions of the breast.

A system that implements some of the concepts presented here has been developed and tested for clinical significance, in order to build a diagnostic decision support system based on the structural features arising from an histologic image of a pathologic duct.

2 The Model of the Duct

Usually the pathologist comes to a diagnosis by viewing at microscope hematoxylin-eosin stained specimens of mammary tissue. The ducts referred by the diagnostic rules of Page appear as in figure 1a: the nuclei are blue, while cytoplasm is pink-red.
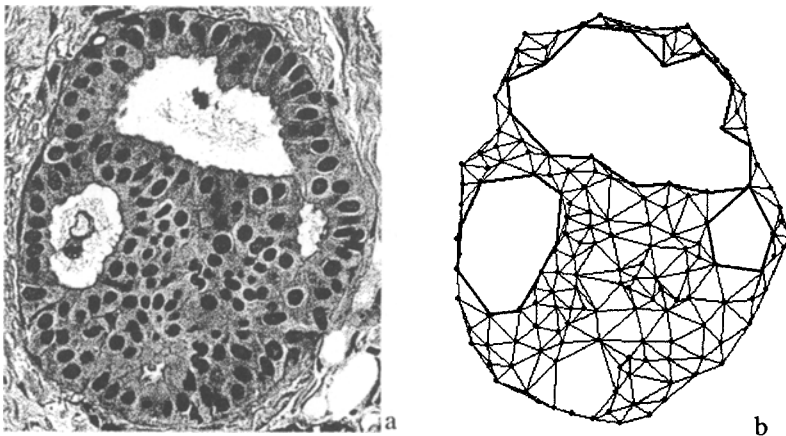

Fig. 1. a) mammary duct, 40x original magnification; b) graph representation

Our technique starts from an histologic image representing the transverse section of a single duct; in fact, the aforementioned rules are referred to a similar image.
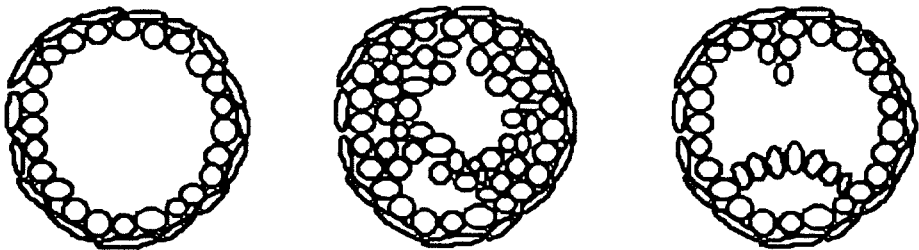

Fig. 2. Ducts

A normal duct is a "pipe" with the wall made by a double layer of cells; the internal layer (epithelial cells) is involved in the lesions (figure 2). The proliferative pathology causes a gradual filling of the internal cavity of the duct, and produces during the intermediate steps some structures, the so-called lumina and bridges; in addition, the filled ("solid") areas show a variable nuclear density. The completeness of the filling

is not relevant for the diagnosis; the fundamental character is the way the proliferating cells occupy the internal space, generating some structures instead of others, or the same structures with different characteristics. In fact, benign cells behave similarly to the starting cells, while the malignant cells usually derive from clonal cells different in shape, dimension and behaviour from the epithelial cells.

The work presented in this paper is an attempt to reproduce the reasoning made by the pathologist when he/she analyzes the spatial arrangement of the nuclei in the duct, as stressed by the approach of Page. For this aim, it is necessary to represent the duct in a structured way so the architectural findings that characterize the different pathologies can be recognized, evaluated, measured and used for the decision support to the diagnosis.

3 The Structure Representation

The primary entities composing the duct are the cells, but the boundaries between them are very difficult to recognize. The cell nuclei are indeed clearly defined, although sometimes they are overlapping, so it is vantageous to use the nuclei as distinctive element for the cell identification. The nucleus has some properties traditionally analyzed with the tools of morphometry and densitometry, as cytoplasmic aspect, nuclear shape, chromasia, but these features - alone - have been demonstrated to be insufficient for a discrimination of the pathologies that have been taken into account.

A relatively new morphometric approach to the study of neoplastic epithelial lesions involves the use of graphs and was described in 1987-1992 by Kayser for lung carcinoma and soft tissue tumours [5-7], in 1992 by van Diest and Co. for breast cancer [12], in 1992 by Meijer and Co. for colorectal adenomatous polyps [8], in 1993 by Raymond and Co. for germinal center of lymph nodes [10], in 1993 by Darro and Co. for colorectal cancer cell lines [3].

These works use mainly the Voronoi tessellation approach, that seems to be suitable to describe space-filling mosaic-like structures resulting from growth processes [4]; also the minimum spanning tree is applied. Both the approaches seem not to be adequate to our need of structuration in the knowledge representation.

The basic idea of our approach is to represent the lesion by means of the neighbourhood relationship between nuclei as primary element that allows the recognition of the spatial properties at a more abstract level. A hierarchical representation based on three main graphs (neighbourhood, planar and dual graphs) is derived from an initial set of nodes that represents the cell nuclei of the lesion, as in the other works [3,5-8,10,12].

A graph G is a mathematical structure defined by a couple of sets N,A:

$$G = <N,A>, N \supset N, A = \{\{n_1,n_2\} \mid n_1 \in N, n_2 \in N\}$$

where N is the node set and A is the arc set; if there exists an arc $a=\{n_1,n_2\} \in A$, we say that n_1,n_2 are connected by a. By means of a semantic function v:

$$v: N \rightarrow D, \forall n \in N \; v(n) = d \text{ , where D is the information domain,}$$

some information can be associated to each node (in our case, the co-ordinates of the nucleus in a local system and eventually some cytological and morphological features of the cell). In the same way it is possible to associate information to the arcs.

We shall consider the neighbourhood graph $G=<N,A>$ $(N \supset N)$ with threshold S defined in the following way:

i) N is the set of natural numbers;

ii) a semantic function v is associated with the node set N;

iii) for each $n \in N$, $v(n)$ represents the two coordinates of n, i.e. $<x,y>$. For convenience, we use two auxiliary functions $x(n)$ and $y(n)$ defined by: $x(n)= x$, $y(n) = y$ $\forall n \in N$, $v(n)=<x,y>$.

iv) $dist(n_1,n_2)$ $\forall <n_1,n_2> \in NxN$ is the usual Ecuclidean distance.

v) the set A consists of all arcs $a=\{n_1,n_2\}$ such that $dist(n_1,n_2) \leq S$.

vi) The length of an arc is defined on A by: $\forall$ $a=\{n_1,n_2\} \in A$, $length(a)=dist(n_1,n_2)$.

The concept of neighbourhood, as expressed by this graph, is redundant because of the great number of intersecting arcs. Having a planar graph is fundamental for our approach because the dual graph, that represents the areas comprised within the arcs, can be generated from it.

The planar neighbourhood graph $P^*=<N^*,A^*>$ may be constructed from G as follows:

i) N^* coincides with N;

ii) A^* is recursively defined by:

$$A_0 = \{a\} \in M_0=\{ m \in A| \ length(m)=\min_{x \in A} length(x)\}$$

$$A_k = A_{k-1} \cup \{a\},$$

$$a \in M_{k-1} = [\ \{ m \in A\backslash A_{k-1} \ | \ length(m) \ = \ \min_{x \in A} length(x)\} \cap$$

$$\{m \in A\backslash A_{k-1} \ | \ \forall \ m' \in A_{k-1} , \ m \ does \ not \ intersect(m,m'))\}]$$

$$A^* = A_k, k = \min_{k \in N} k : A_k=A_{k+1}$$

In this way, the complexity of the representation decreases, and from the planar graph another step can be made to obtain the dual graph. In the dual graph each node represents an area (called "face") comprised among arcs of the dual graph, and remains associated with the arcs of the planar graph that form the perimeter of the face. The last two graphs provide a core for the structural analysis of the lesion, adopting techniques as previously described elsewhere [13,14]. Each graph may be partitioned into subgraphs representing some structures of the lesion, or using criteria related to cytologic uniformity [7].

The main subgraphs represent respectively the lumina, the solid area and the external perimeter of the duct. The lumina are defined as faces of the dual graph with a relatively high number of sides and such that the corresponding region in the starting image contains a white area; the perimeter of the duct is the external face of the dual graph; the solid area comprehends the remaining faces.

These subgraphs may be analyzed in order to quantify the features described by Page or eventually discover other criteria; for example, the subgraph representing the solid area may give useful data related to the dimensions and regularity of the cells analyzing the distribution of the arc lengths or of the face areas, while the roundness and cribriformity of the lumina may be measured on the lumina subgraph by using some adequate shape factors.

This structural representation may be included in a decision support system able to generate a diagnostic suggestion starting directly from an histologic image. In such a

system, it is necessary to study four modules: a perception module that realizes the segmentation tasks, a module for the generation and evaluation of the structural representation, a module for the management of the diagnostic knowledge (learning and use), and a man/machine interface.

4 Results

A prototype of the system for the structural analysis of lesions has been developed (SANE, Structural ANalysis Environment) [1]. At this stage of development, the prototype comprehends the perception module, the representation module capable to recognize only some structural features and a simple decision-support module based on decision trees. The algorithms for the derivation of the planar and dual graphs have been studied and tested.

We focused mainly on structure representation, so the perception module is actually implemented with a simple segmentation technique using filtering, thresholds and mathematical morphology procedures. For each region recognised as a nucleus, center of gravity, area, major axis and shape factor are computed and saved as attributes for the corresponding node.

For the preliminary analysis of the system, we studied images obtained from 70 cases. Each image represents a breast duct from hematoxylin-eosin stained specimens, acquired at 40x magnification by a videocamera connected to a motorized microscope, and processed by the system. Each image measured from 512x512 to 1536x1536 pixels, 24 bit depth. Only a limited subset of graph features was taken into account.

As an example, the planar/dual graph for the duct of figure 1a is shown in figure 1b.

To obtain a first evaluation of the clinical significance of this approach, the statistical analysis of the differences in some features of the ducts was carried out. Among those considered, we mention mainly the statistical characterization of solid area, obtained by using mean, standard deviation, minimum and maximum of node degree and arc lengths on neighbourhood and planar graphs, and of perimeters and areas for dual graph. The analysis of these features revealed some significant differences about cell dimensions and variability among the three diagnostic classes (table 1), validating our approach to the structure representation of the lesions as previously shown in [1]. In addition, the significant features were submitted to an automatic generator of decision trees; the results of this are yet to be evaluated.

feature	description	EPI	ATHYP	CIS	Tukey-Kramer HSD p value
GAMAX	Neighbourhood graph: max node degree	29.83±4.2	22.21±4.2	15.29±4.1	0.001
GAMED	Neighbourhood graph: mean node degree	16.48±2.8	12.56±1.8	7.90±2.4	0.001
GASD	Neighbourhood graph: SD node degree	5.252±1.0	3.874±0.8	2.851±0.7	0.01
PLMED	Planar graph: mean arcs length	40.00±3.2	44.43±2.9	51.79±4.7	0.05
PERSOL	Solid dual graph: mean face perimeter	119.3±9.4	132.7±9.0	154.5±14	0.05

Table 1. Significant features

5 Conclusions

The preliminary results obtained by the prototype system are encouraging. The possibilities of this technique are to be explored more deeply, with particular attention to the features not yet considered. Moreover, the knowledge based module that realizes the decision support to the diagnosis should be improved with the management of uncertainty.

This approach seems also to be applicable to other similar classification problems in the field of histopathology.

References

1. C.A. Beltrami, V. Della Mea, N. Finato: Structure Analysis of Breast Lesions using Neighbourhood Graphs. Anal Quant Cytol Histol, 17, 143-148 (1995)
2. College of American Pathologists, Consensus Meeting: Is "fibrocystic disease" of the breast precancerous? Arch Pathol Lab Med 110,171-173 (1986)
3. F. Darro, A. Krukzynski, C. Etievant, J. Martinez, J.L. Pasteels, R. Kiss: Characterization of the differentiation of human colorectal cancer cell lines by means of Voronoi diagrams. Cytometry 14, 783-792 (1993)
4. U. Hahn, U. Lorz: Stereological analysis of the spatial Poisson-Voronoi tesselation. J Microsc 175, 176-185 (1994)
5. K. Kayser, H. Stute: Minimum spanning tree, Voronoi's tesselation and Johnson-Mehl diagrams in human lung carcinoma. Path Res Pract 185, 729-734 (1989)
6. K. Kayser, K. Sandau, G. Bohm, K.D. Kunze, J. Paul: Analysis of soft tissue tumors by an attributed minimum spanning tree. Anal Quant Cytol Histol 13, 329-334 (1991)
7. K. Kayser, K. Sandau, J. Paul, G. Weisse: An approach based on two-dimensional graph theory for structural cluster detection and its histopathological application. J Microsc 165, 281-288 (1992)
8. G.A. Meijer, P.J. Van Diest, J.C. Fleege, J.A. Baak: Syntactic structure analysis of the arrangement of nuclei in dysplastic epithelium of colorectal adenomatous polyps. Anal Quant Cytol Histol 14, 491-498 (1992)
9. D.L. Page: Cancer risk assessment in benign breast biopsies. Hum Pathol 17, 871-874 (1986)
10. E. Raymond, M. Raphael, M. Grimaud, L. Vincent, J.L. Binet, F. Meyer: Germinal center analysis with the tools of mathematical morphology on graphs. Cytometry 14, 848-861 (1993)
11. S.J. Schnitt, J.L. Connolly, F.A. Tavassoli, R.E. Fechner, R.L. Kempson, R. Gelman, D.L. Page: Interobserver reproducibility in the diagnosis of ductal proliferative breast lesions using standardized criteria. Am J Surg Pathol 16, 1133-1143 (1992)
12. P.J. Van Diest, J.C. Fleege, J.A. Baak: Syntactic structure analysis in invasive breast cancer: analysis of reproducibility, biologic background, and prognostic value. Hum Pathol 23, 876-883 (1992)
13. L. Vincent: Graphs and Mathematical Morphology. Signal Processing 16, 365-388 (1989)
14. C.T. Zahn: Graph-Theoretical Methods for Detecting and Describing Gestalt Clusters. IEEE Trans. on Comp. c-20/1, 68-86 (1971)