Pattern Recognition Letters 7 (1988) 245–252
North-Holland

April 1988

Characterization and featuring of histological section images

B.B. CHAUDHURI*, K. RODENACKER and G. BURGER

Gesellschaft für Strahlen- und Umweltforschung mbH München, Institut für Strahlenschutz, D-8042 Neuherberg, Fed. Rep. Germany

Received 18 December 1986
Revised 23 January 1988

Abstract: Image analysis of the structural organization of cells in histological sections has been shown to be useful for the quantitative characterization of tissue. The organization may be expressed in the form of a tree or a graph structure. Features derived from the graph along with the cell features can be used effectively for case classification and diagnosis. The paper describes techniques of generating the tree and graph as well as deriving features from them. Results of applying the techniques to cervical tissue section specimen images are also presented.

Key words: Histology, histometry, image analysis, segmentation, feature extraction, pattern recognition.

1. Introduction

In cyto- and histopathology, cellular specimens and histological sections are investigated to search for pathological changes in cell and tissue morphology. In many cases the diagnostic goal is not only a binary qualitative decision, but also the grading of a continuous changing situation. This can be assisted by applying quantitative imaging methods establishing new possibilities in analytical and quantitative cytometry and histometry.

The standard approach in image cytometry is to segment the cell images in the specimens and derive useful features for the single cell classification. Specimen classification is normally a two-stage approach where the individual cells are classified at first and then the results are aggregated by some rule to obtain the class status of the specimen (Burger and Jütting, 1986). The local distribution of the cells in common cytological specimens is generally considered to be less important for classification and hence ignored in most cases for featuring. In the histological section images, on the other hand, the arrangement of the cells provides a very important clue for the diagnosis of pathological tissue growth and differentiation.

One of the arrangement properties is the topology. It may be characterized by defining object neighbourhoods. A graph can then be constructed where the cell nuclei denote the graph vertices and the neighbourhoods the graph edges. From the graph, features for the characterization of the tissue may be derived. The present paper describes the general steps of the histometric procedure.

2. General featuring concepts

It is convenient to classify the characteristic features of histological sections into individual cell features and features representing the local arrangement of the cells in the histological section. The latter may be called 'topological' features (Rodenacker, 1987b).

* Guest Scientist from: Electronics and Communication Sciences Unit, Indian Statistical Institute, Calcutta, India.

0167-8655/88/$3.50 © 1988, Elsevier Science Publishers B.V. (North-Holland)

2.1. Cell level features

Cell level features are not directly related to the goal of this paper and will only be mentioned briefly. One can distinguish several types of cell features that are useful for characterization. Besides color, these are (a) morphological features, such as the projected area or perimeter, shape and orientation features, etc. (Preston, 1981). The relative coordinates of the center of gravity of the cells may also be treated as cell features. (b) photometric features such as the integrated optical density (IOD) and (c) textural features, describing the staining pattern of the cell (Rodenacker, 1987a).

Using ILIAD procedures (Eriksson et al., 1982) more than 150 cell level features are derived at this laboratory for the analysis of smear specimens. Most of them are not directly applicable for fractions of objects appearing in sections. We have chosen a few for our purpose. They are all related to the segmented cell nuclei. These are the area and perimeter, IOD, center of gravity coordinates, barycentric coordinates and the angular orientation of the line of best fit through the center of gravity, called the major axis direction. At the next section we shall describe how they are useful for the characterization of the tissue topological features.

2.2. Topological features

One of the possible ways of representing the local arrangement of objects is the use of 'graph' or 'tree' structures. Informally speaking, a graph is a set of 'vertices' and 'edges' where an edge represent the connection between two vertices according to certain property. We can think of traversing from one vertex to the other through a concatenated set of edges called 'path'. A path is a 'loop' if it is closed. A tree or 'spanning tree' is a graph without any loop and is connected i.e., having a path from any vertex to any other. If the vertices are defined over a metric space then an edge may be characterized by its length which is the distance between corresponding vertices. A 'minimum spanning tree' is a tree where the sum of the lengths of the edges in the tree is minimum.

To represent the histological section image as a graph or tree structure, the segmented cell nuclei (nuclear profiles) are used to represent the vertices, which are located at the 'centers of gravity'. In addition the vertices carry the cell features as labels or markers.

The most common topological property used in choosing the edges is the neighbourhood of cells. One possible way to define neighbourhood is the Voronoi tesselation (Voronoi, 1902) of the section area considering the graph vertices as a marked random point process (Stoyan and Mecke, 1983). With the Voronoi tesselation each vertex is enclosed by a polygon. The neighboured vertices are those which have a common polygon side. To generate the graph all neighbouring vertices are connected by edges. Prewitt (1979) and Sanfeliu et al. (1981) used this approach for muscle tissue analysis. An alternative definition of neighbourhood due to O'Callaghan (1975) is based on two constraints namely the distance constraint and the direction constraint. A modified version of this neighbourhood is used by Kayser and Hoffgen (1984) to generate graphs for the analysis of colon mucosa. In this paper we describe two other techniques of graph generation. They are based on the nearest neighbourhood and the zone of influence neighbourhood.

From the generated graph or tree a set of features is derived. The total graph (or tree) may also be dissected into subgraphs (or subtrees) which are then subject to featuring. Derived features may include the number of edges per vertices, the number of loops and also higher statistics from frequency distributions of edge and vertex features of certain subsets.

3. Minimum spanning tree

In a Euclidean space the minimum spanning tree (MST) is sequentially formed as follows. Let V denote the set of vertices. Start at any vertex in V and connect it by an edge with the vertex nearest to it. Include these vertices in a set, say U. Now, at each stage connect a vertex in U with another vertex in $V - U$ for which the distance is the minimum of all such pairwise distances. Include the newly connected vertex in U and repeat the process until $V - U$ is empty. The MST has some interesting

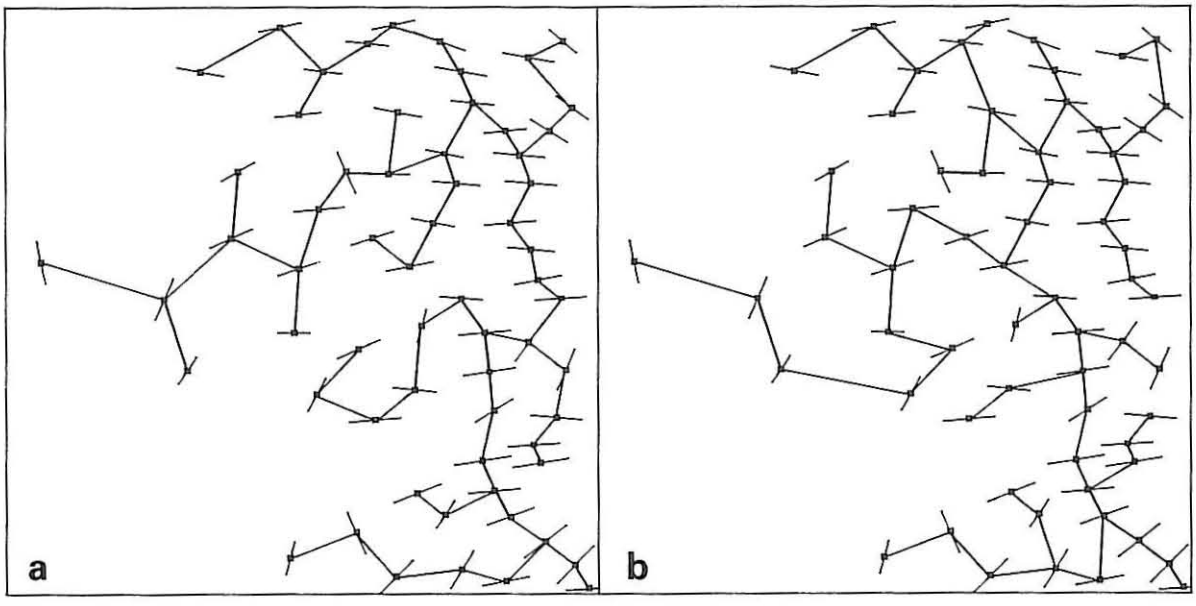

Figure 1. (a) Minimum spanning tree corresponding to Figure 4(a); (b) with orientation weighted distance. (Direction of orientation of each nuclear profile is also shown.)

properties. One of them is that its generation does not depend on any parameter other than the starting vertex. Also, the subtrees of a MST are again MST. In Figure 1(a) a MST is shown where the centers of gravity of the nuclear profiles are chosen for the vertices. It corresponds to the microscopical field of a histological section from the ectocervix, shown in Figure 4(a).

We have also done a modification of the MST where the distance between the vertices is weighted by a factor that depends on the relative orientation of the nuclear profiles at the vertices. Let θ be the smaller of the angles between the major axes of two cells. If d denotes the Euclidean distance between the cells than the modified distance d' is defined as

$$d' = d(1 + \sin(\theta)). \tag{1}$$

Now the minimum spanning tree is constructed using d' for finding the nearest neighbours. The idea is to connect primarily those cells with each other which are more aligned and therefore more likely to belong the same generation layer of cells. As an example, the minimum spanning tree with the modified distance for the Figure 4(a) is shown in Figure 1(b).

4. Zone of influence tesselation graph

Consider the two-tone segmented image consisting of the cell nuclei and their background. The zone of influence (ZOI) tesselation is formed by the skeletonization or medial axis transformation of the background image. There exists a wide variety of algorithms for the skeletonization of a two-tone image. The basic idea is to blow cells in a grassfire fashion. Where the borders of neighbouring growing cells meet, skeleton pels are defined. If the background image is topologically simply connected then the skeletonization results in the partitions of spaces each of which contains a cell nucleus (zone of influence ZOI). The algorithm used here is the sequential application of the so-called hit or miss transformation widely used in mathematical morphology (Serra, 1982). The result of tesselation for Figure 4(a) by this approach is shown in Figure 2(a).

From Figure 2(a) it is seen that the border of ZOI of a cell is common to the border of ZOI of other cells which we call neighbouring cells. A graph is generated if the neighbouring cell vertices are connected by edges. As in the case of minimum spanning tree we can use the center of gravity as the cell

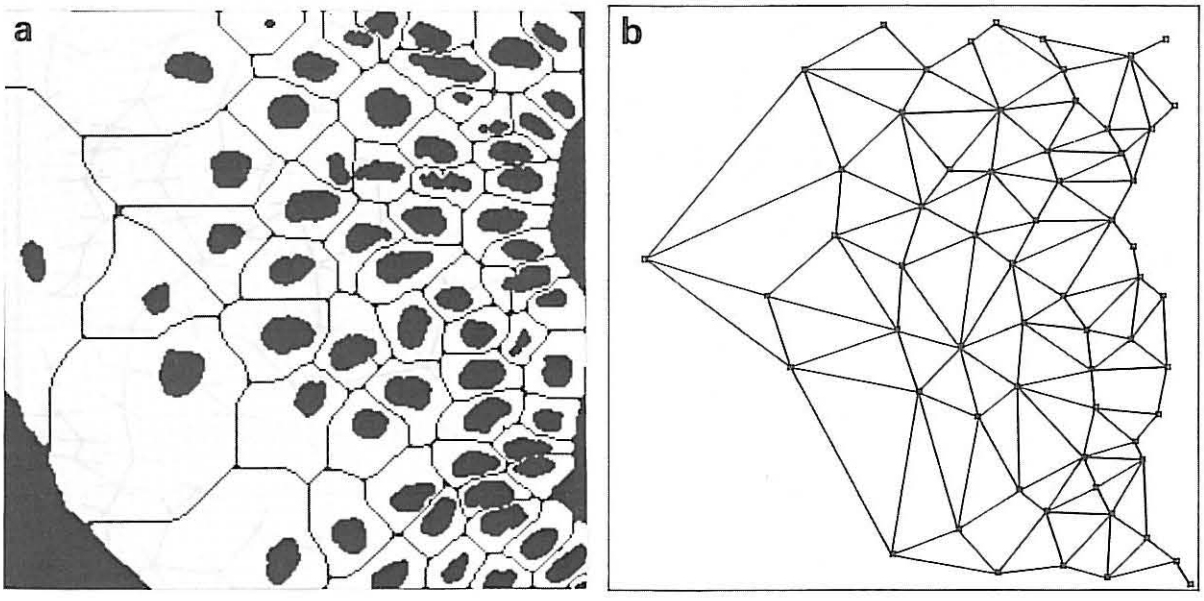

Figure 2. (a) Zone of influence tesselation; (b) Corresponding ZOI graph.

vertex. The resultant graph structure for Figure 2(a) is shown in Figur 2(b).

There are some advantages of generating neighbourhood graphs by this method as compared to those defined by O'Callaghan's method and the Voronoi diagram. In O'Callaghan's method it is necessary to fix some parameters such as maximum possible distance and the minimum angular separation of the neighbours with respect to the candidate point, which is not necessary in this method. Also, both O'Callaghan's method and the Voronoi diagram work on points and no information about the shape of the cell nucleus is utilized in defining the neighbourhood. On the other hand, shape information is important in ZOI tesselation. However, the neighbourhood defined by this method would be similar to the Voronoi neighbourhood if the cells are represented as points and circular propagation is used for skeletonization and this approach may be termed as generalized Voronoi tesselation (Arcelli and Samiti di Baja, 1986).

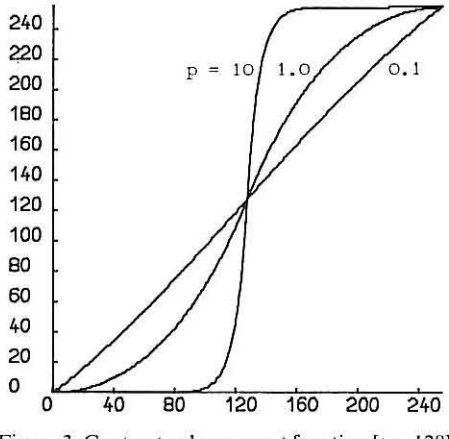

Figure 3. Contrast enhancement function [$t = 128$].

5. Clinical application

5.1. Material and method

A set of cervical tissue section images are taken in order to test and find the useful and important features for characterization. The sections are obtained from conisations and they are usually 4 to 8 microns thick. They are stained with haematoxilin and the digital image is obtained with the BASISS system (Gais et al., 1984) by this laboratory.

Image aquisition is performed by a TV-scanning microscope with 25 × or 50 × objective, using a narrow band optical filter of 546 nm average wavelength. The field sizes are 500 × 500 micron^2 or

250 × 250 micron^2 and digitized, according to the German TV-norm, into 512 × 512 pixels. The local resolution is then nominally 1 micron or 0.5 micron pixel distance.

The spatial resolution chosen is a compromise on cell and tissue region size. In the examples shown the pixel size is 0.5 micron. The image to be processed contains 256 × 256 pixels only (with 256 gray levels). The pathological diagnosis of the region for this investigation are divided in two groups namely, carcinoma in situ (CIS) and normal (N).

5.2. *Preprocessing and segmentation*

To characterize a histological section image it is first necessary to segment the cell nuclei from the background. Because of various factors of preparation and imaging there is a considerable variability of gray levels in the cells and the background. The main problems in normal sections are caused by the occluding cells distributed in depth. Use of local and global segmentation techniques (Rosenfeld and Kak, 1976) which are available in standard software packages (SPIDER, 1983) does not yield satisfactory results and necessitates a high degree of succeeding manual interaction. To reduce this, we adopted a two-stage procedure.

The first stage is contrast enhancement by the use of a nonlinear gray scaling that is dependent on local properties of the image. Let the image gray levels range between 0 and 255. The transformed gray level is given in terms of the input gray level I by the function.

$$I' = \mathrm{Int}\left[I\left(\tan \frac{\pi}{4} \frac{I}{t} \right)^p \right], \quad I \le t \tag{2}$$

and

$$I' = 255 - \mathrm{Int}\left[(255 - I)\left(\tan \frac{\pi}{4} \frac{255 - I}{255 - t} \right)^p \right], \; I > t, \tag{3}$$

where $0 < t < 255$ is a threshold, p a positive real quantity and $\mathrm{Int}(I)$ means integer part of I. From eq. (2) it is clear that for $I = t$, $I' = I$ i.e., gray level t remains unchanged in the transformed image. The transformation is obviously enhancing contrast, controlled by the parameter p. A typical plot of the function for a few values of p is shown in Figure 3. For our images it is found that satisfactory results are obtained if t is chosen at a histogram valley. The initial estimate of p is changed according to the average gray value of the neighbourhood. p is defined as

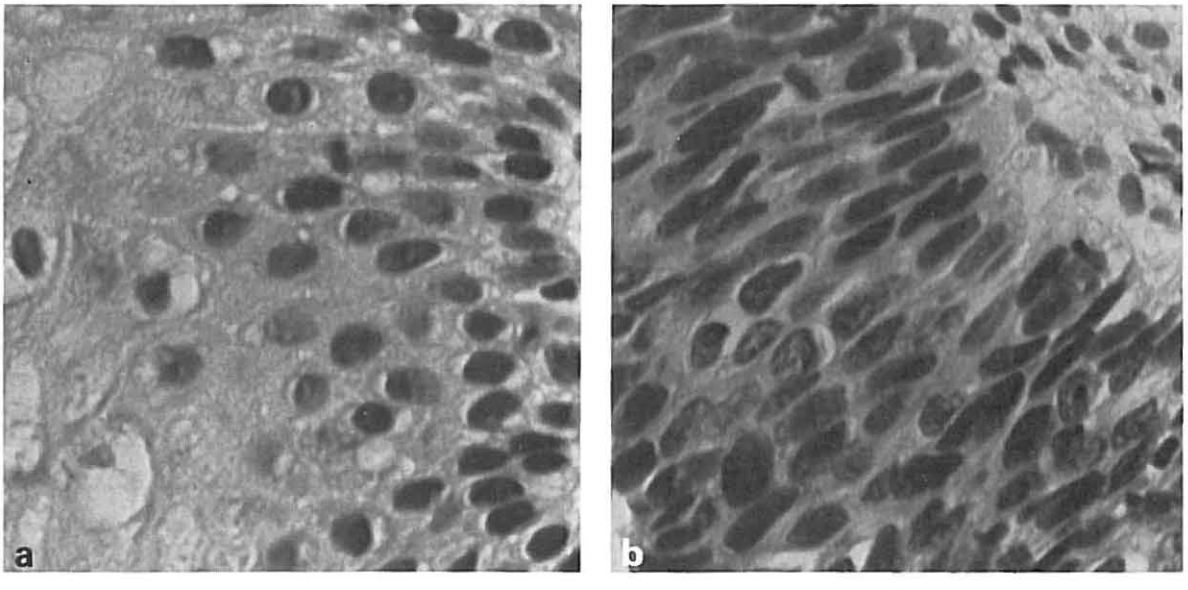

Figure 4. Normal and abnormal regions of cervical tissue sections. (a) Normal; (b) Carcinoma in situ (CIS).

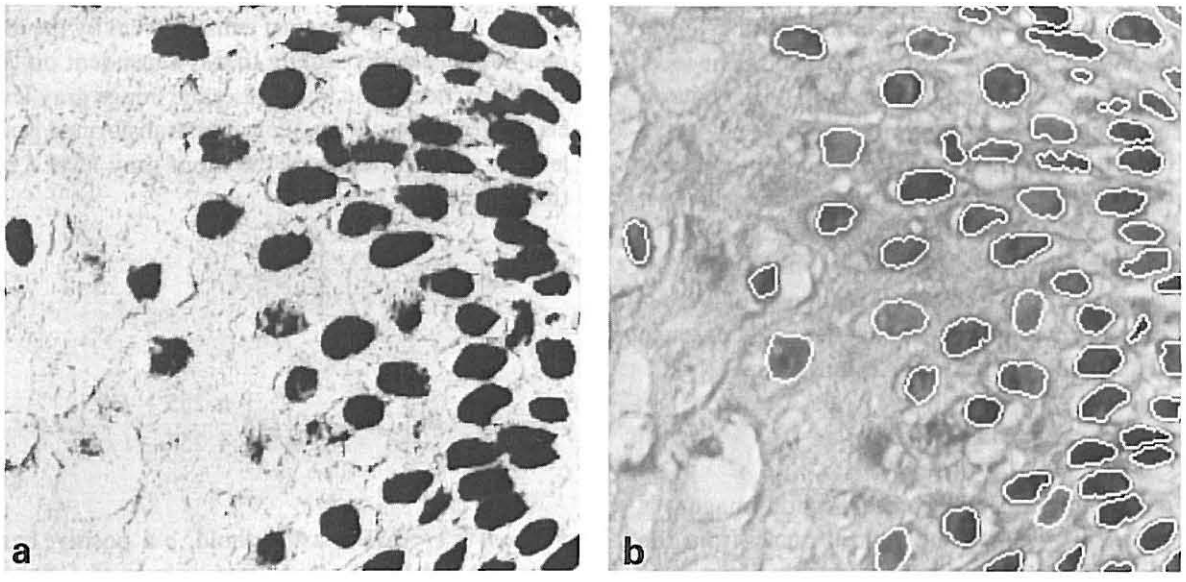

Figure 5. (a) Enhanced image of Figure 4(a); (b) Segmented nuclei border superposed on Figure 4(a).

$$p = 10/(\text{magnitude}\ (I - I_{3\times 3}) + 1) \qquad (4)$$

where $I_{3\times 3}$ denotes the average gray value of the three by three neighbourhood and I the gray value of the candidate pixel.

For the second stage, the segmentation, a threshold is chosen from the histogram valley of the resultant image. Hole filling, noise cleaning and smoothing are done by opening and closing operations of mathematical morphology (Serra, 1982) and nonlinear median filtering (Huang et al., 1978). Finally, manual interaction is employed to separate the remaining occluding objects, the extent of which depends on the specimen. Two typical histological section regions are shown in Figure 4 while the enhanced and segmented version of one of them is shown in Figure 5.

5.3. Results

Some global features have been derived from the graph and tree structures as follows: Let d_i denote the average length of the edges connected to i-th nucleus. We may consider d_i as a local density measure of cells in the section. The frequency distribution $f(d_i)$ obviously displays the distribution of cells in the section. The first and second moment of $f(d_i)$ (mean M and standard deviation S of $f(d_i)$) are then global features for 'cellular density'. The same holds for the coefficient of variation $R = S/M$. Typical values of these features for both MST and the tesselation graph are given in Table 1.

In order to extract more information than the first two moments from d_i one may generate suitable histograms by partitioning the whole available range of d_i into a few intervals. The histograms may then be evaluated by means of vector or correspondence analysis respectively. Table 2 gives some indication on the different distribution for 3 bin-histograms with n_1, n_2 and n_3 representing the histogram values.

Table 1

	Minimum spanning tree			ZOI tesselation graph		
	M	S	R	M	S	R
Normal 1	21.95	7.71	0.351	30.03	12.11	0.403
Normal 2	24.29	9.36	0.385	32.49	12.16	0.374
CIS 1	16.17	4.10	0.253	22.64	5.36	0.236
CIS 2	21.96	3.69	0.168	29.72	5.50	0.185

Table 2

	n_1	n_2	n_3
Normal 1	52	8	3
Normal 2	37	24	6
CIS 1	32	84	10
CIS 2	12	45	35

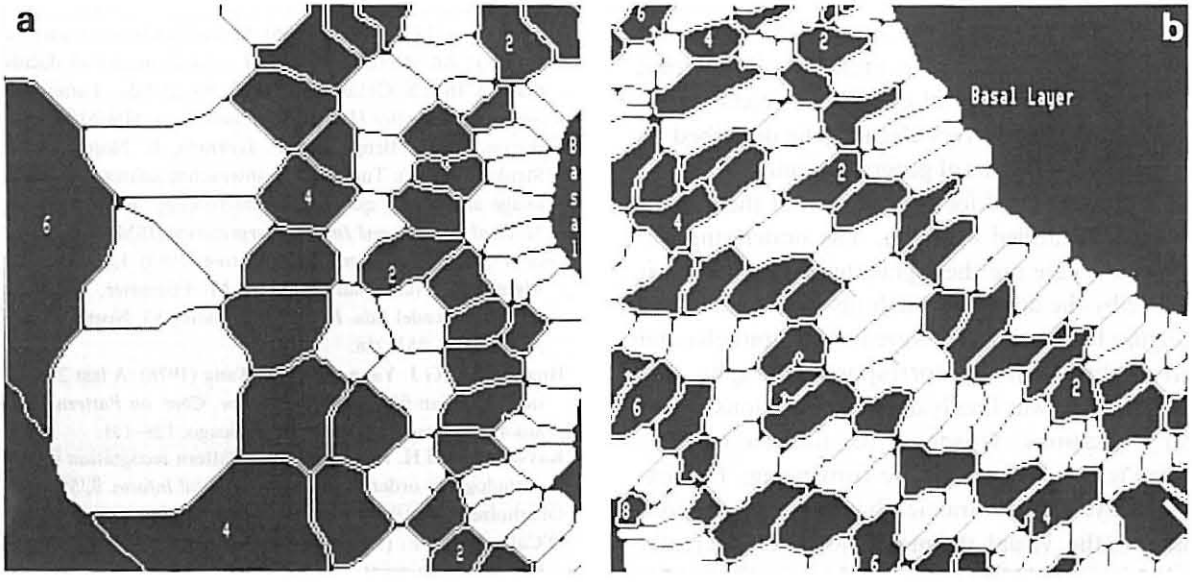

Figure 6. (a) Growth layers corresponding to the region of Figure 4(a); (b) Growth layers corresponding to the region of Figure 4(b). (Even layers numbered.)

Simple classification is not always the goal of histological section image analysis. In this context, especially for intraepithelial changes in squamous epithelium, it may be of interest to find the growth direction and growth pattern of the cell layers in the section. We can use the ZOI neighbourhood to visualize the growth pattern. In this technique, an end layer say, the basal layer, is interactively marked as an initial set of nuclei. The algorithm now detects ZOI neighbours of the successive layers of growth. A typical layering by this approach is illustrated in Figure 6 that corresponds to the images in Figure 4. Properties of these layers may be used for characterization of the section especially when it is necessary to distinguish different grades of neoplasia such as mild, moderate and severe dysplasias. Often the pathologist visually partitions the section in three nearly equal regions between the basal and the superficial layers and examines the density, orientation and shape of the cells in each of the regions to distinguish different stages of malignancy (Oberholzer, 1986).

The layering shown in Figure 6 may be used to quantify the approaches of the pathologist. Figure 7 shows an example for the above mentioned layer approach for normal (n), light (l), moderate (m) and severe (s) dysplasia and carcinoma in situ (c) specimens. Features used are the area of ZOI per layer. The slope of the regression line per section reflects the dynamics of changes of cell distribution from basal layer to the section surface.

6. Summary

Histometry is a field of growing interest in quantitative and analytical cytopathology. Its goals ran-

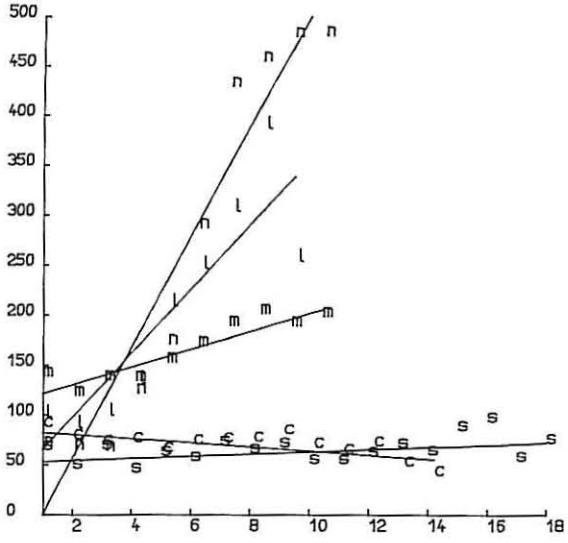

Figure 7. Mean area of ZOI per layer with regression lines for normal (n), light (l), moderate (m) and severe (s) dysplasias and carcinoma in situ (c) specimens. (Images are taken with 25 × objective.)

ge from measurement of the amount of tissue components to the topometric analysis of the local arrangement of structural entities or objects of interest. The latter can very elegantly be described by graphs. Several means of generating suitable graphs have been presented for the analysis of the growth pattern of stratified epithelia. The underlying objects in this case are the segmented cell nuclei. It is possible by the described methods to automatically recognize the layered structure in such epithelia and derive features from the corresponding graphs.

The method was finally applied to sections of cervical conisations. Whenever the number of cases studied is low, the results are convincing. The features derived show drastic changes in correspondence to the visual diagnosis, for the differently graded intraepithelial dysplastic lesions. They offer the quantitative continous grading of such lesions in preoperative biopsies and may help to establish a more stringent level of therapeutical intervention then hitherto.

Acknowledgement

The authors are grateful to Dr. M. Oberholzer, Kantonsspital Basel, Switzerland and Prof. W. Gössner, from the Pathology Dept. of GSF and Technical University Munich for histological specimens and many valuable discussions. One of the authors (B.B. Chaudhuri) wishes to thank the International Büro, GSF, for support under Indo-German Scientific Co-operation and the Director of the Indian Statistical Institute for approving the sabbatical leave to conduct the work. Interest in the work by Prof. D. Dutta Majumder is also acknowledged with thanks.

References

Arcelli, C. and G. Samiti di Baja (1986). Computing Voronoi diagrams in digital pictures. *Pattern Recognition Letters* 4, 383–389.

Burger, G. and U. Jütting (1986). Specimen classification in cytometry: An intercomparison of various means of decision making. In E.S. Gelsema and L.N. Kanal Eds., *Pattern Recognition in Practice II*. Elsevier, Amsterdam, 509–519.

Eriksson, O., E. Bengtsson, T. Jarkrans, B. Nordin and B. Stenkvist (1982). Tuning of an interactive software system for image analysis to quantitative microscopy. *Proc. Int. Symp. Medical Imaging and Image Interpretation* (ISMII 82), Berlin.

Gais, P., K. Rodenacker and U. Jütting (1984). Lokale logische Algorithmen für Binärbilder. In: M. Feiemeier, G. Joubert and U. Schendel Eds. *Parallel Computing* 83. North-Holland, Amsterdam, 253–258.

Huang, T.S., G.J. Yang and Y.G. Tang (1978). A fast 2-dimensional median filter algorithm. *Proc. Conf. on Pattern Recognition and Image Enhancement*, Chicago, 128–131.

Kayser, K. and H. Höffgen (1984). Pattern recognition in histopathology by orders of textures. *Medical Inform.* 9, 55–59.

Oberholzer, M. (1986). Private communication.

O'Callaghan, J.F. (1975). An alternative definition for 'Neighbourhood of a point'. *IEEE Trans. Computer*. 24, 1121–1125.

Preston, K. (1981). Tissue section analysis: Feature selection and image Processing. *Pattern Recognition* 13, 17–36.

Prewitt, J.M.S. (1979). Graphs and grammers for histology: An introduction. Proc. Third Annual Symp. Comp. Appl. and Medical Care, Washington.

Rodenacker, K. (1987a). Featuring of cellular objects. In: G. Burger, J.S. Ploem and K. Goertler, Eds. *Clinical Cytometry and Histometry*. Academic Press, London, 91–96.

Rodenacker, K., P. Bischoff and B.B. Chaudhuri (1987b). Featuring of topological characteristics in digital images. *ActaStereol* 6, 945–950.

Rosenfeld, A. and A.C. Kak (1976). *Digital Picture Processing*. Academic Press, New York.

Sanfeliu, A., K.S. Fu and J.M.S. Prewitt (1981). An application of a distance measure between graphs to the analysis of muscle tissue pattern recognition. Saragota Springs, New York, 86–89.

Serra, J. (1982). *Image Analysis and Mathematical Morphology*. Academic Press, London.

SPIDER User's Manual (1983). Joint System Development Corporation, Tokyo.

Stoyan, D., J. Mecke (1983) *Stochastische Geometrie*. Akademie, Berlin.

Voronoi, G. (1902). Nouvelles application des parametres continus a la theorie des formes quadratiques. Deuxième Memoire: recherches sur les parallelloedres primitifs. *J. Reine Angew. Math.* 134, 198–287.