Densitometric features of cell nuclei for grading bladder carcinoma

(세포핵 조밀도에 의한 방광암의 진행 단계)

Heung-Kook Choi* and Ewert Bengtsson**

*Dept. of Biomedical Engineering, Medical Research Center, Seoul National University

**Centre for Image Analysis, Uppsala University, Uppsala, Sweden

Abstract

A way of quantitatively describing the tissue architecture we have investigated when developing a computer program for malignancy grading of transitional cell bladder carcinoma. The minimum spanning trees, MST was created by connecting the center points of the nuclei in the tissue section image. These nuclei were found by thresholding the image at an automatically determined threshold followed by a connected component labeling and a watershed algorithm for separation of overlapping nuclei. Clusters were defined in the MST by thresholding the edge lengths. For these clusters geometric and densitometric features were measures. These features were compared by multivariate statistical methods to the subjective grading by the pathologists and the resulting correspondence was 85% on a material of 40 samples.

Introduction

Quantitative, computer assisted, analysis of tumor tissues started in the late sixties. Using computerized image analysis, many of features that are qualitatively used by pathologists can be quantitatively described. Additional features are not readily apparent to the human eye such as high order texture descriptors. Since much of the information used to grade the tissue comes from the relations between the cells, i.e. an assessment of the order vs. disorder of the tissue organizatoin, it is interesting to try to describe these relations. Prewitt proposed to use graph analysis as mathematical tools for this [11]. Since then several other authors have followed up on the idea and developed graph analysis algorithms for tissue characterization [6,8,12]. In particular the MST has proven useful. A few years ago we started a project the aim of which was to develop an easy-to-use, reproducible, objective, computerized grading system for bladder tumors. The reproducibility of subjective grading seems to indicate considerable prognostic power [3]. In the project we have previously studied parameters obtained after segmentation of the cell nuclei [7], as well as texture parameters [4]. Other groups have presented similar results [10]. In the work reported here we extend this work by investigating the use of the MST to describe the tissue architecture.

Material and Methods

Material and image acquisition

All patients with newly detected transitional cell carcinoma for the urinary bladder seen at Uppsala University Hospital during the period

1975-1978, were studied retrospectively with an observation time ranging from 5 to 9 years. Our subjective grading divided the material into four groups: grade 1, 2A, 2B and 3 . Grade 1 has a slight papillary tendency (Fig. 1a). In grade 2A the cellular pattern shows more variation in nuclear size, shape, internuclear distance and chromatine features (Fig. 1b). In grade 2B the general impression of disorder dominates (Fig. 1c). Grade 3 exhibits an extreme variation in each of the features described (Fig. 1d).

A standard monochrome video CCD camera, giving a data matrix of 512X410 pixels was used on a microscope with an interference filter centered at 550 nm and 40X lens, making the size of the pixels about 0.5 μm. Digitized images were obtained from 5 μm thick stained paraffin wax sections. Pathologists chose a region of interest from each section which best represented the subjective grade. This image acquisition and processing was carried out on an EPSILON workstation using EGO software system [1].

Histogram based thresholding

The gray scale image was automatically segmented into a binary image with objects (nuclei) and background. The method global thresholding, was based on the fact that cell nuclei appear as dark objects on a lighter background. A histogram of a typical image showed a dominant distribution of light pixels, the background, and a smaller distribution of darker pixels, the objects. The two distributions were separated by s weak valley or an inflection point. We found the threshold by searching for minima in the first derivative of the histogram.

Watershed segmentation

The result binary image showed the nuclei but also many other spurious objects, holes, as well as touching and overlapping nuclei. The object masks were first cleaned filling the holes of the objects, defined as having an area less than 70 pixels or a maximum width less than 9 pixels, were detected. The width was obtained by a chamfer 3-4 distance transformation [2]. The distance transform image (Fig. 2a) was also used as a starting point for separating touching objects. The watershed segmentation algorithm [13] can separate irregular blobs into more regular parts. We empirically determined that a waistline between blobs with a width of less than 90% of the smallest side blob should result in a cut. For this cut the watershed line was replaced with the shortest straight line, see Fig. 2b.

Minimum spanning tree

A minimum spanning tree (MST) connecting the centroids of all nuclei in a tissue section can be used as a basis for extracting features describing the tissue architecture. We first created a linear graph connecting the centre points of adjacents. The edges were assigned the Euclidean distance between the corresponding vertices as weights. This graph was then reduced to a MST using Kruskal's algorithm [9]. In Kruskal's algorithm, edges are first sorted in ascending order of weight and placed in a list structure. This algorithm was implemented here using a heapsorted priority queue. As the list is traversed, if an edge is found connecting a vertex that has not been visited before, the edge is included in

the tree. When an edge connects two vertices belonging to different clusters, the two clusters are merged, with the new cluster being absorbed into the older clusters. The tree will complete when only one cluster remains i.e. when all the vertices have been visited. For N vertices, the complete tree will have (N-1) edges. The obtained tree will be the MST i.e. the graph whose weightsum is minimum among all spanning tree graphs. A practical example is shown in Fig. 2c.

Clustering

In the MST the vertices (nuclei) were clustered based on cutting links longer than a threshold value. Reasonable value for this threshold were obtained by studying the image geometry. Two slightly overlapping nuclei would have an edge distance of about 6 μm and a distance of more than 19 μm is rather unusual. Several distance in this range were used as possible thresholds for creating clusters. The threshold which yielded features with the best discriminating power was finally used. Fig. 2d. depicts the clusters of segmented nuclei.

Feature extraction

We defined 24 different features. The best of these as determined by their discriminatory power were: The degree of nuclear overlap defined as the number of edges shorter than 6 μm divided by the total number of nuclei (F1), the variation of local orientation of the graph defined as the standard deviation of the direction indices measured in a 2X2 (F2) or 3X3 (F3) neighborhood, the nuclear area to graph size

ratio defined as the sum of all nuclear areas divided by the total graph size (F4), the normalized number of clusters defined as the number of clusters divided by the total graph size (F5), the variation in average nuclear size defined as the standard deviation of the average nuclear areas in each cluster, taken over all clusters (F6) and similarly the standard deviation for the average gray values of the nuclei (F7), the variation in inter cluster distance as the standard deviation of the distance between the clusters (F8) and finally the cluster area variation defined as the standard deviation of total nuclear area per cluster over all clusters (F9). These features achieved discriminating powers for the different classification phases as shown in Table 1.

Hierarchical classification

A two stage binary hierarchical classifier using multivariate linear discriminant analysis at each stage was developed. For each stage we applied a step-wise linear discriminant analysis, using the BMDP [5] program package. The optimized criteria are the ratio between the within group and between group distance . using the Mahalanobis distance. For each phase in the hierarchical classification the program found the optimal combination for features and the corresponding weights in the discriminant function. The discrimination between grades (1 + 2A) and (2B + 3) used clusters with less than 16μm distance (Phase I) while the threshold 7 μm and 4.5 μm were used for grade 1 and 2A (Phase II:1) and grade 2B and 3 (Phase II:2) respectively.

Result and Discussion

The classifier was trained on a randomly selected subset of 20 out of the 40 images and then tested on all 40 images, 10 images of each grade. The classifier was able to correctly classified 34/40 images (85 % agreement) see Table 2. The method thus showed good correlation with subjective visual grading. This at least indicates that there is some useful information available in these features. These preliminary results need to be verified on a larger test set before we finally can know if these methods can be used as part of such a system. Computing the MST is a computationally demanding task. We used Kruskal's algorithm with s time complexity of $O(n \log n)$, where e is the number of edges in the graph.

Acknowledgments

References

[1] Bengtsson E, Eriksson O: The design of an image analysis system. 5th Scan Conf Image Analysis, 217-224 (1987).

[2] Borgefors G: Distance transforms in digital images. Comput Vision Image Process 34, 344-371 (1986).

[3] Busch C., Engberg A., Norlen B.J., Stenkvist B.: Malignancy grading of epithelial bladder tumours. Scan J Urol Nephrol 11, 143-148 (1977).

[4] Choi H.K., Vasko j., Bengtsson E., Jarkrans T., Malmstroem P.U., Wester K. & Busch C. Grading of transitional cell bladder carcinoma by texture analysis of histological sections. Anal Cell Pathol 6, 327-343 (1994).

[5] Dixon W J.: BMDP Statistic software manual, 7M. Depart. Of biomathematics, University of California Press. Berkeley, Los Angeles, Ocford, 339-358 (1990).

[6] Dussert C., Rasigni M., Palmari J., Rasigni G., Llebaria A., Marty A.: Minimum spanning tree analysis of biological structures. J Theor Biol 125. 317-323 (1987).

[7] Jarkrans T., Vasko J., Bengtsson E., Choi H.K., Malmstroem P-U., Wester K., Busch C.: Grading of transitional cell bladder carcinoma by image analysis of histological sections. Anal Cell Pathol 8, 135-158 (1995).

[8] Kayer K., Stute H., Tacke M.: Minimum spanning tree, integrated optical density and lymph node metastasis in bronchial carcinoma. Analyt Cell Pathol 5, 225-234 (1993).

[9] Kruskal J.B.: On the shortest spanning subtree of a graph and the traveling salesman problem. Proc Amer Math Soc 7, 48-50 (1956).

[10] de Meester U., Young I.T., Lindeman J., van der Linden H.C.: Towards a quantitative grading of bladder tumors. Cytometry 12, 602-613 (1991.

[11] Prewitt J.M.S.: Canonical representation for tissues and textures. IEEE computer Society: COMPSAC, 470-476, (1979).

[12] Rodenacker M., Aubele M., Jutting B., Gais P., Burger G., Gossner W., Oberholzer M.: Image cytometry in histological sections. Acta Stereol 11, 249-254 (1992).

[13] Vincent L., Soille P.: Watershed in digital space: An efficient algorithm based on immersion simulations. IEEE Trans on PAMI 13, 583-598 (1991).

Fig. 1. The digital microscopic images (40X) of the bladder carcinoma cells stained by Feulgen. In each grade the region of interest was chosen by pathologists: (a) Grade 1(upper left), (b) Grade 2A(upper right), (c) Grade 2B(lower left) and (d) Grade 3(lower right).
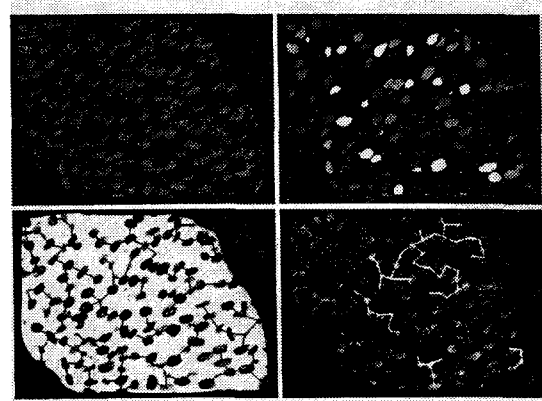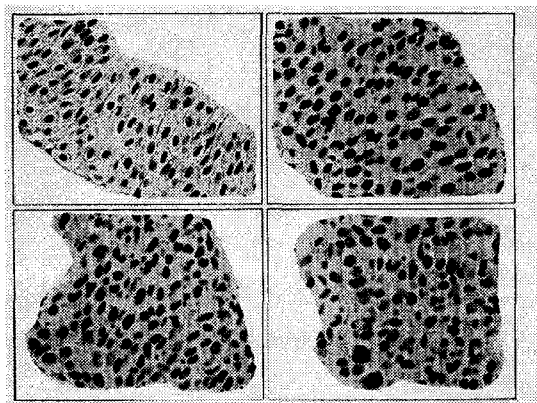

Fig. 2. The grade 2A image of Fig. 1b. (a) Upper left, using a chamber 4-3 distance transformation . (b) Upper right, the overlapping nuclei separated by a watershed algorithm. (c) Lower left, the minimum spanning tree of a tissue section with 153 vertices superimposed on the gray scale image. (d) Lower right, a thresholded at 16 μm has been used as the Euclidean distance between vertices to create the 31 clusters.

Table 2. Confusion matrix for the minimum spanning tree based hierarchical classifier tested on 40 images. The kappa value was 0.80 and the t-value was 8.83. Visual grade horizontally, computer grade vertically.

Grade	Grade 1	Grade 2A	Grade 2B	Grade 3	%
1(10)	8	2	0	0	80
2A (10)	0	9	1	0	90
2B (10)	0	1	9	0	90
3(10)	0	1	1	8	80
Total (40)	8	13	11	8	85

Table 1. The following features were selected for each of the three classification stages and training phase using 20 randomly selected images. The corresponding means, standard deviations, variation ranges and the F-values for each of the groups are given.

Features	Mean	Std	Min	Max	Mean	Std	Min	Max	F-dist
Phase	Grade 1, 2A				Grade 2B, 3				
F3	14.54	2.31	10.33	18.54	10.95	4.51	2.25	17.1	7.77
F4	11.43	1.94	8.66	14.52	14.93	2.15	12.64	19.17	12.81
F6	61.59	38.23	0	105.76	240.95	128.49	83.99	459.75	5.76
Phase II:1	Grade 1				Grade 2A				
F1	0.012	0.01	0	0.025	0.007	0.008	0	0.02	34.11
F3	15.07	2.5	12.1	18.54	14.01	2.26	10.33	16.49	8.35
F5	0.042	0.004	0.034	0.045	0.042	0.004	0.037	0.047	15.76
F7	14636	2357	11409	16675	16983	4572	10596	22742	44.09
F9	97.77	9.38	86.25	107.91	126.98	26.98	83.52	153.75	92.69
Phase II:2	Grade 2B				Grade 3				
F2	13.63	3.74	8.12	17.66	10.48	4.51	3.92	16.37	4.43
F4	15.98	2.23	14.07	19.17	13.87	1.63	12.64	16.24	9.41
F6	180.2	39.28	150.7	246.5	224.3	70.44	165.3	325.8	7.64
F8	0.445	0.093	0.354	0.596	0.513	0.186	0.298	0.782	6.25