H.W. Schüssler (editor)
Elsevier Science Publishers B.V. (North-Holland)

DETECTION OF TUMOR CELLS IN THE URINE

C.-E. Liedtke, Ph.D.; B. Aeikens, M.D.

Universität Hannover, Medizinische Hochschule Hannover
Hannover, Federal Republic of Germany

The aim of the project is to find out to which extent tumor cells can be separated from normal cells by high resolution image analysis techniques. For the quantitative analysis of cells new reliable and robust algorithms for image segmentation have been developed. Features have been extracted from the cells which give special consideration to nuclear structure. By applying this feature set to our present data base of 1239 cells a correct classification of approximately 90% has been obtained for the differentiation between normal and tumor cells.

1. INTRODUCTION

It is the goal of an interdisciplinary project funded by the Ministry of Research and Technology of the Federal Republic of Germany to develop new and practical methods for the early detection of bladder cancer and for use in follow-up studies of patients after treatment, mainly transurethral resection. One aim of this project is to find out to which extent the different types of malignant urinary cells can be separated from each other and from normal cells by high resolution image analysis techniques.

The cells are obtained by filtering of voided urine with a 5µ-membrane filter and are then attached to a microscopy slide. For staining we use Papanicolaou stain and apply the staining procedure, which has been earlier described by Wied. The cells are sampled and digitized at 7 bit amplitude and 0.3 µm spatial resolution by use of a photometer microscope at 54o nm wave length. The cells are classified into tumor, normal and atypical cells by visual inspection. Our data set consists now of 1239 cells from approximately 2o patients of both sexes.

2. SEGMENTATION

The segmentation of a microscopic cell scene into cell nucleus, cytoplasm, background and artefacts is an easy task for an human observer but turns our to be extremely difficult and error prone for a machine. The reason that humans perform so much superior to machines is that they use in addition to the pictorial information, which becomes visible through the microscope, a-priori-knowledge about those specific cell scenes - the type of information which we usually describe by "experience". A machine segmentation will become the more accurate, reliable and insensitive to small changes in cell preparation, the more a-priori-knowledge about the specific cell scene can be incorporated in the segmentation algorithm. We favour a segmentation method which we call "blob detection and assembling" method. This method is based on the fact that in agreement with human perception segments, which are by our definition meaningful subimages like "nucleus", "cytoplasm", "cell", etc. either consist of one or are composed of several uniform regions which we call "blobs". Therefore our strategy which is shown in the block diagram of Figure 1 is first to find the blobs and then assemble them to segments by use of prior knowledge about the content of the image. The application of the algorithm to cell scenes is demonstrated in Fig.2. The final result including other cell scenes are presented in Figures 2G to 2I.

We have put much emphasis and much computational effort into segmentation because errors in segmentation may propagate into the following step of quantitative image analysis which is feature extraction.

3. FEATURE EXTRACTION

From several discussions we expected that the relevant information about malignancy of cells is contained in the structure of the nucleus. Therefore we tried to describe nuclear structure parametrically. The method is illustrated in Fig.3. The spatial gray level distribution of a nucleus (Fig.3A) is decomposed by thresholding at equidis-

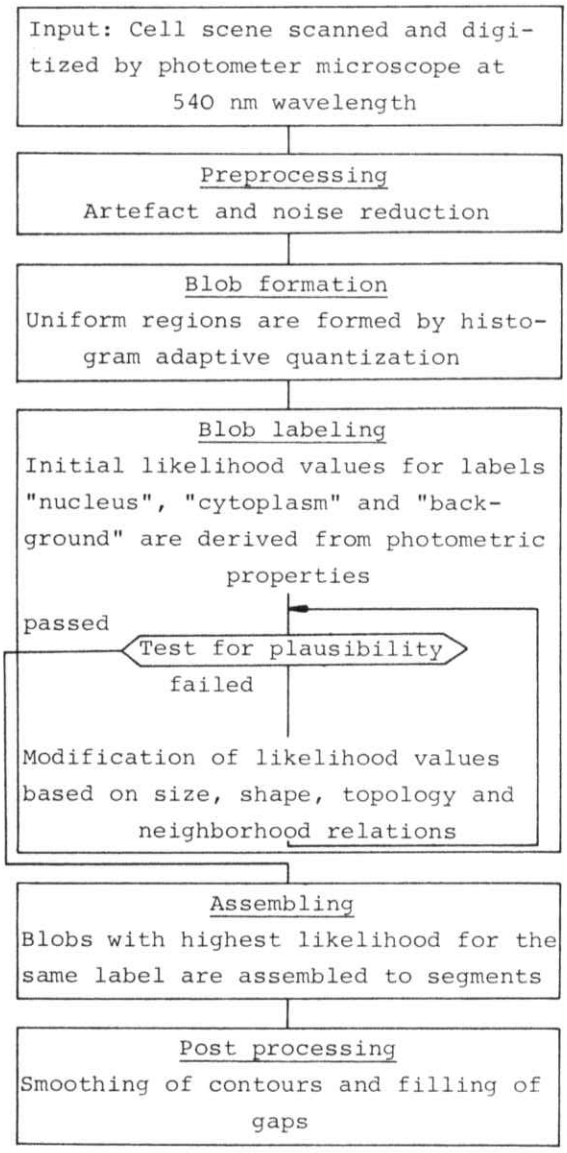

Fig.1: Block diagram of the blob detection and assembling method for image segmentation

tant gray levels into slices (Figures 3B - 3I). Each slice consists of connected components of different size. By ordering these components from left to right according to decreasing area and under preservation of the topological relations between neighboring slices the 3D-representation of Fig.3A is reduced to a 2D-representation of Fig.3J. The peaks in Fig.3J from left to right represent a base-structure and sub-structures of different order from the original image together with their respective gray levels. For a parametric description of the structure various measurements have been derived from the 2D-representation like the "volume" of the base-structure, the number of sub-structures, maximum, minimum and range of the gray levels in the base- and substructures, etc. The feature set has been augmented by several measurements describing photometric and geometric properties of the cytoplasm and relations between respective nuclear and cytoplasmic measurements.

4. RESULTS

This feature set has been tested on our learning set of 1239 cells. The evaluation was based on a reclassification which yields usually optimistic results, which are better than the performance of the system and a 10%-jackknive-test, which gives usually pessimistic results, i.e. classification results, which are worse than the true performance of our classification method.

The following results have been obtained by using a multiple linear regression method. The classes are labelled with N for "normal", T for "tumor" and A for "atypical" cells.

Reclassification with 10 features

visual class. \ computer classification	N	T	A
N	94,3%	2,0%	3,7%
T	12,5%	84,6%	2,8%
A	50,2%	40,2%	9,6%

10%-jackknive test with 10 features

visual class. \ computer classification	N	T	A
N	93,7%	2,0%	4,3%
T	12,5%	83,9%	3,5%
A	51,4%	40,6%	8,0%

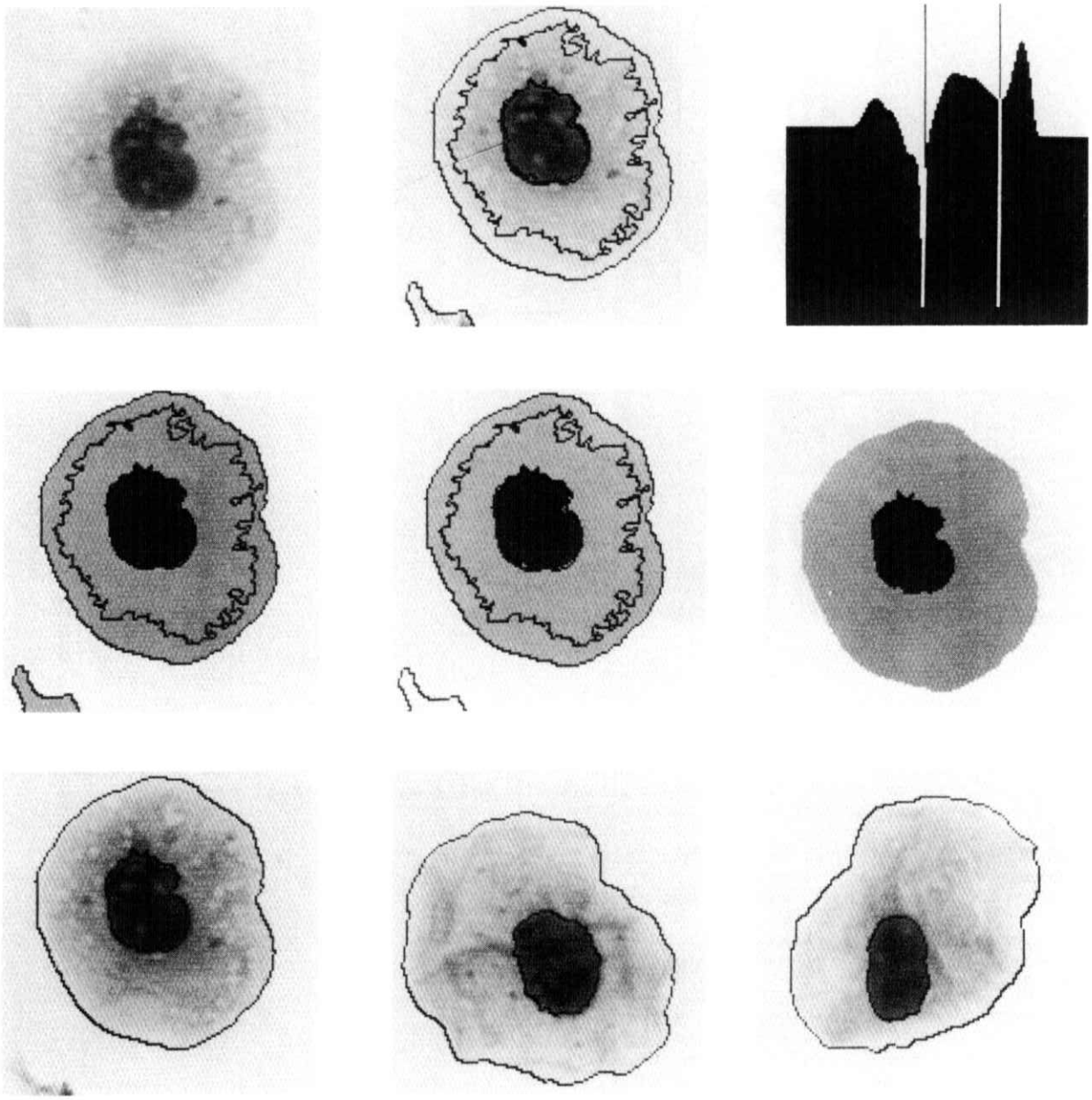

Fig.2: Blob detection and assembling algorithm:
(A) original scene, (B) blob detection by histogram adaptive thresholding, (C) thresholds in the histogram indicating gray level regions for nucleus, cytoplasm and background, (D) blob labeling based on photometric values: nucleus = black, cytoplasm = gray, background = white, (E) correction of labels based on topology and neighborhood relations, (F) assembling of blobs with the same label to segment masks, (G) segment borderlines overlaid to original scene, (H,I) results for two other cells

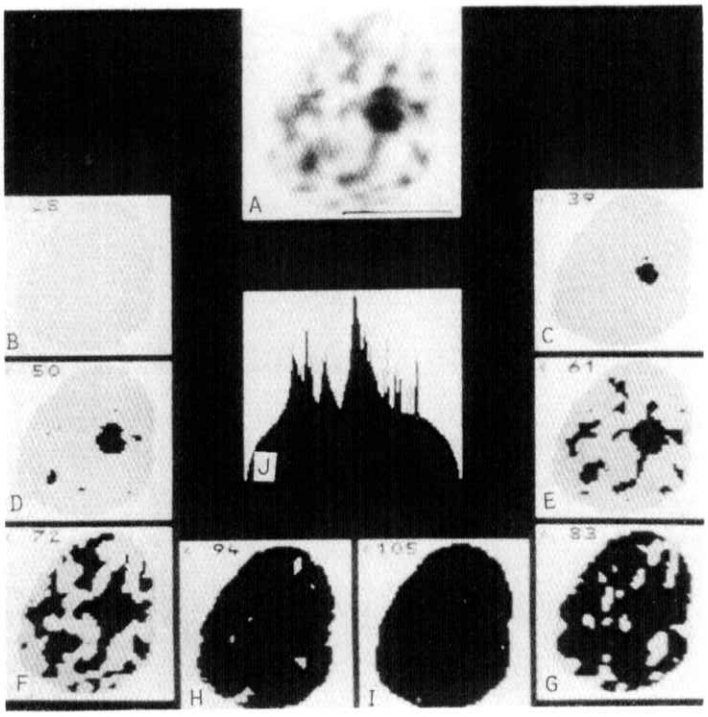

Fig.3: Transformation of 3D representation (gray level vs 2D-space) of a cell nucleus (A) into 2D representation (gray level vs size) (J) via binary threshold masks (B-I)

The "best features for the differentiation of normal, tumor and atypical cells have been found to be
- Average contrast within the cytoplasm
- Minimum transmission within the nucleus
- Ratio between nuclear and cytoplasmic area
- Average transmission within the cytoplasm
- Average transmission within the nucleus

Two important conclusions could be drawn from these and several other tests on our data:

1. There is no indication that atypical cells form a separate cluster, equivalent to normal and tumor cells in the feature space. It is more likely that atypical cells belong to the clusters of normal and tumor cells but are located in the border region between normal and tumor cells in the feature space.

2. The relevant information for the differentiation between normal and tumor cells is neither restricted to nuclear structure nor to other nuclear measurements as has been assumed before. Measurements from the cytoplasm are at least equivalent to those from the nucleus in their discriminatory power for separating single normal from tumor cells.

For the two class problem normal vs tumor cells we obtained approximately 9o% correct classification.