Texture analysis of cervical cell nuclei by segmentation of chromatin patterns

Smeulders, A.W.M.; Leyte-Veldstra, L.; Ploem, J.S.; Cornelisse, C.JJ.

DOI
10.1177/27.1.374575

Publication date
1979

Published in
The Journal of Histochemistry and Cytochemistry

UvA Keur
UB Groningen
Broerstraat 4
9700 AN Groningen

A078327016 ISN: 681391 RAPDOC (R) **OPGEHAALD** NCC/IBL

Verzoeke te behandelen voor: 27-04-2005 **Ingediend door:** 0004/9999
Datum en tijd van indienen: 13-04-2005 20:23 **Datum plaatsen:** 13-04-2005 20:23 **Type aanvrager:** UKB **I.D.:** UVA KEUR (UB GRONINGEN)

PPN: 840044240

journal of histochemistry and cytochemistry Histochemical Society 1953 Baltimore, MD Williams & Wilkins

Gewenst: 1979-00-00 **Deel:** 27 **Nummer:** 1 **Electronisch leveren (LH=N)**

Auteur:	**Titel van artikel:**	**Pagina's:**
A.W.M. Smeulders, L. Leyte-Veldstra, J.S. Ploem, C.JJ. Cornelisse	Texture analysis of cervical cell nuclei by segmentation of chromatin pa	199-203

Opmerking:
arno ID: 136042

ZODIAC	MAG	NN39135	Vol. 21(1973)-
WWW			Vol. 45(1997)-

1. ○ origineel gestuurd
2. ○ fotokopie gestuurd
3. ○ overige
4. ○ nog niet aanwezig
5. ○ niet aanwezig
6. ○ niet beschikbaar
7. ○ uitgeleend
8. ○ wordt niet uitgeleend
9. ○ bibliografisch onjuist
10. ○ bij de binder

Fakturen zenden aan: Rijksuniversiteit Groningen
Bibliotheek, Uitleenbureau
Postbus 559
9700AN Groningen

0022-1554/79/2701-0199$02.00/0
THE JOURNAL OF HISTOCHEMISTRY AND CYTOCHEMISTRY

Vol. 27, No. 1, pp. 199-203, 1979
Printed in U.S.A.

Texture Analysis of Cervical Cell Nuclei by Segmentation of Chromatin Patterns

A. W. M. SMEULDERS, L. LEYTE-VELDSTRA, J. S. PLOEM AND C. J. CORNELISSE

Department of Pathology (A.W.M.S.; C.J.C.) and Department of Histochemistry and Cytochemistry (L.L.-V.; J.S.P.) University of Leiden, Leiden, The Netherlands

Received for publication July 19, 1978.

Texture parameters of the nuclear chromatin pattern can contribute to the automated classification of specimens on the basis of single cell analysis in cervical cytology. Current texture parameters are abstract and therefore hamper understanding. In this paper texture parameters are described that can be derived from the chromatin pattern after segmentation of the chromatin pattern. The image segmentation procedure is based on a region grow algorithm which specifically isolates high chromatin density. The texture analysis method has been tested on a data set of images of 112 cervical nuclei on photographic negatives digitized with a step size of 0.125 μm. The preliminary results of a classification trial indicate that these visually interpretable parameters have promising discriminatory power for the distinction between negative and positive specimens.

In cervical cytology automation the evaluation of chromatin patterns by texture parameters can contribute to the classification of specimens on the basis of single cell analysis. Computationally efficient Markovian (8) and Laplacian (1) texture parameters appear to have high discriminatory power for the distinction between normal and abnormal chromatin patterns but their performance can be evaluated only by using a classification algorithm since these abstract parameters are difficult to interpret visually. The validity of this evaluation strongly depends on the reliability of the data base used for testing of these parameters which is subject to the ambiguity in visual pattern classification by the cytologist. Parameters which more closely describe what the cytologist sees may facilitate a better understanding of the textural pattern types, in addition to the evaluation with the aid of a classification algorithm.

In this study we explored the possibility that such texture parameters may be obtained by segmentation of the visually conspicuous chromatin regions with a relative high optical density (O.D.). The segmentation of these regions requires an algorithm that operates on local O.D. maxima in the digitized image rather than using a fixed threshold level only (5). With the aid of an O.D. threshold regions which are separately perceived may emerge into one region in the digitized image.

We developed an algorithm which permits the consecutive segmentation of regions with relative high optical densities in the digitized image starting from local O.D. maxima. The performance of the algorithm and the discriminatory power of the texture parameters derived from the segmented images have been tested on a data set of 112 nuclear images from normal and abnormal cervical cells digitized from photographic negatives at a high gray level and spatial resolution. The preliminary results indicate that image segmentation can yield visually perceivable texture parameters that can contribute to the discrimination between negative and positive cervical specimens in an automated classification procedure.

Data acquisition: For testing the performance of the developed texture analysis technique a data set was composed of cells from three cytologic classes. Apart from negative specimens, two opposite positive specimen classes were included (three negative specimens, eight moderate to severe dysplasia and four invasive carcinoma). In this preliminary study cells from carcinoma *in situ* specimens were omitted, since this intermediate diagnostic class would have introduced too much uncertainty in the evaluation.

From the positive specimens only nuclei with a DNA content over 5 C have been selected for the data set. This has been found to be a suitable prescreening criterion for the identification of positive specimens in a study of 122 negative and 334 positive cases by Ploem-Zaaier *et al.*[1] The nuclei selected according to this criterion show a variety of chromatin patterns. In negative specimens nuclei with a DNA content over 5 C are found only rarely and therefore initially nuclei from normal intermediate cells have been included in the data set from the negative specimens.

To permit the selection of nuclei on the basis of their DNA content the specimens were stained with the acriflavine-Feulgen-stilbene (AFS) procedure for DNA and protein (2). The morphologic identification of AFS stained cells is performed on the fluorescence image excitated with transmitted violet light. For the microfluorometric determination of the DNA content the acriflavine fluorescence is stimulated with incident blue excitation light. Transmitted blue light illumination (466

[1] Ploem-Zaaier JJ, Beyer-Boon ME, Leyte-Veldstra L, Ploem JS: Cytofluorometric and cytophotometric DNA measurements in cervical smears stained with a new bicolor method. To be published in the Proc. of the Second International Conference on Automated Cytology and Image Analysis, Tokyo Japan 1976.

nm) is used for photography of the acriflavine absorption image which presents the chromatin pattern very well. The visual selection, microfluorometry and photography of nuclei is carried out on a Leitz MPVII microfluorometer (E. Leitz Inc., Rockleigh, N. J.) equipped with an Orthomat W camera according to the procedure described by Ploem-Zaaier *et al.*[1] For photography a film was chosen with a contrast index (slope of the Hurter-Driffield curve) of about 0.7, since this matches optimally the maximum optical densities found in AFS stained nuclei ($\simeq$ 1.3) to the dynamic range of about one decade of the AD convertor in the microdensitometer used. This condition is met by Kodak Panatomic X (Eastman Kodak Co., Rochester, N. Y.) film developed in D 76 (Kodak, diluted 1:1 with water) for 3 min at about 20°C without stopbath. For four different Panatomic X films the contrast index varied from 0.61 to 0.70 which is quite acceptable taking into account the variation in staining intensity among different slides (9).

For the present investigation a small spatial and gray level digitization step is used, in order to be able to choose an optimal resolution for a real time system. The spatial resolution of the microscope system used is theoretically about 0.23 μm as has been confirmed by previous work at our laboratory.[2] From the work of Groen (4) on chromosome analysis on the same microscope system, it appeared that the 8th bit of a gray level carried information. Therefore the photographic negatives are embedded between glass slides (11) and scanned with a Leitz MPVI microphotometer equipped with a 400 Hz scanning stage with 40 μm step size developed in the Department of Histochemistry and Cytochemistry. In this way the nuclear images are digitized with a final step size of 0.125 μm at the level of the nucleus. The photomultiplier signal is converted by an 8-bit AD convertor. Scanning is controlled by the HIDACSYS software package (10) running on a PDP 11/10 minicomputer (Digital Equipment Corp., Maynard, Mass.) which is connected to a PDP 11/60 for further processing of the data.

Image processing: For the analysis of the digitized images the software package TIC (texture of images of cells) has been developed, written in FORTRAN IV and for a small part in PDP assembler. In the first image processing step the module DIODA (10) isolates the nucleus in the digitized image and computes global parameters such as nuclear area, integrated O.D. and average O.D. The second step involves the segmentation of the regions with relative high optical densities in the image starting from local O.D. maxima (starting points) which are found by scanning the digitized image from top left to bottom right. The segmentation procedure is restricted to a fixed percentage of the nuclear area by an O.D. threshold computed for each nucleus individually from the intranuclear histogram. Only picture points with O.D. values above this threshold are included in the segmentation procedure. The individual segments are found by dilatation from the starting points to adjacent points with continuously decreasing O.D. values. Dilatation stops when the threshold O.D. value is reached or when the slope in O.D. is no longer declining. In the present investigation the segmentation is heuristically restricted by the O.D. threshold to 20% of the nuclear area. This represents approximately the most conspicuous chromatin regions.

A line printer representation of the partitioning of the digitized image of the nucleus depicted in Figure 1 is given in Figure 2. A good correspondence with the regions of high chromatin density in the original image is apparent.

For each segment the following properties are determined: (*a*) peak value (mean of the 9 highest O.D. values); (*b*) area; (*c*) integrated O.D.; (*d*) coordinates of the O.D. peak; (*e*) length of the borderline(s) with other segment(s).

From this set of properties for each segment a number of texture parameters can be derived for the complete nuclear image (Table I). The first six parameters of Table I have a rather global character whereas parameters No. 7–10 are indicative for the regularity of the nuclear pattern with respect to the O.D., the distribution and the degree of clumping of the highly condensed regions. For parameter No. 6 the second highest local O.D. maximum has been taken instead of the first maximum to avoid the possibility that the presence of a Barr body would reduce the discriminatory potential of this parameter.

Classification: Evaluation of all kinds of parameters based on the properties of the segments would require a larger data set than the 112 nuclear images available at present. The classification trial is therefore restricted to a test on the discriminatorial potentials of six texture parameters. The texture parameters selected describe the staining intensity and spacing of the condensed chromatin regions in the nuclei. The statistical evaluation of the data set is performed with the aid

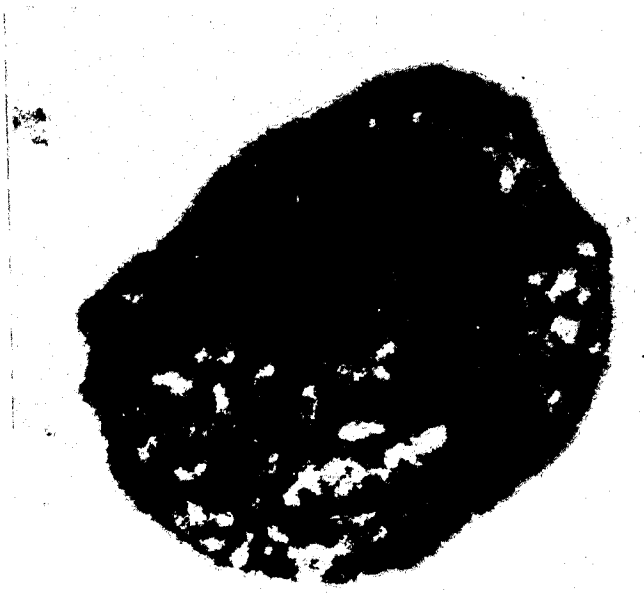

FIG. 1. Abnormal nucleus from a cervical cell stained with acriflavine-Feulgen-stilbene. Absorption image photographed at 466 nm on Kodak Panatomic X film. Magnification approximately ×4000.

[2] Van den Berg H: Op het scherp van het beeld, M.Sc. Thesis, Delft, 1974 (in Dutch).

FIG. 2. Line printer representation of the segmented regions with relative high chromatin contrast found by the region grow algorithm in the digitized image of the nucleus shown in Figure 1. The area of the individual segments is marked by a character indicating the sequence in which the segments are found. The vertical enlongation of the picture is caused by the line printer format.

TABLE I

Texture Parameters Derived from the Properties of the Individual Segments Per Nucleus

1. Number of segments (TOT)
2. Total area of all segments (TOTARE)
3. Integrated O.D. of all segments (TOTCON)
4. Mean area of all segments (MEAARE)
5. Mean O.D. of all segments (MEACON)
6. (*a*) Second highest local O.D. maximum (DARK), (*b*) normalized with respect to the mean O.D. of the nucleus (DARKN)
7. Sum of the squared differences between the local O.D. maximal values ranked according to their O.D. (DIFF)
8. Second order statistics of the coordinates of the segment points (STAT2)
9. Mean length of the borderline between the segments (TOUCH)
10. Number of free lying segments (FREE)

of the software package ISPAHAN developed by Gelsema (3) for interactive statistical evaluation of data.

The mean values of all six texture parameters are listed in Table II and apart from FREE, show a good individual distinction between the three cytologic classes. *t*-Tests against a level of significance of 1% show that the mean value of five parameters differ significantly for the three classes (see Table III).

The following ranking in discriminatory power for these six texture parameters is found (the names refer to Table I): 1, MEACON; 2, DIFF; 3, DARK; 4, FREE; 5, TOUCH; 6, DARKN.

As stated above at this time cells from carcinoma *in situ* specimens were omitted from the analysis. For the classification trial these six texture parameters are used. The data set is divided in two parts using the first part as a training set for the classification of the second part and *vice versa*. The combined results are compared with the cytologic classification in Table IVA. A classification is also performed using the training set as test set (Table IVB). Comparison of the Tables IV A and B shows a good reproducibility of the classification result. The confusion matrix in Table IV A and B shows that the nuclear images from negative specimens were sometimes assigned to the dysplasias. Nuclear images from the dysplasia *a priori* class are for the major part assigned to the same *a posteriori* class whereas a minor part is assigned to the negative and invasive carcinoma class respectively. The majority of the nuclear images in the invasive carcinoma *a priori* class are also classified correctly with some overlap with the other two classes.

DISCUSSION

The use of the chromatin pattern in automated cell classification is complicated by several factors operating at different steps in the specimen handling, data acquisition and data processing procedures. A first complication is that the type of pattern obtained is dependent to a certain extent on the fixation and staining method used. Uncontrolled variations in these procedures may artificially increase the variability in patterns. In this respect a well defined cytochemical staining procedure for DNA as used in this study can be of advantage since it stains specifically a unique molecular component of chromatin.

A second complication is found at the data acquisition level. To retain the information present in the detailed chromatin pattern digitized images are needed that have a high spatial and gray level resolution. In the present study this was approximated by using an intermediate photographic enlargement step and 8-bit AD convertor to achieve an image digitized with a step size of 0.125 μm.

Although this step size provides sufficient image points to perform image segmentation operations on detailed structures

TABLE II

Mean Values and Standard Deviations of Six Texture Parameters for Nuclei from Three Different Cytologic Specimen Classes

Parameter[a]		Class		
		Negative	Dysplasia	Invasive Carcinoma
MEACON	[a.u.]	0.88 ± 0.24	1.43 ± 0.33	1.98 ± 0.73
DIFF	[a.u.]	2.94 ± 2.79	0.76 ± 1.26	0.20 ± 0.33
DARK	[a.u.]	1.23 ± 0.25	1.82 ± 0.28	2.21 ± 0.68
FREE	—	45 ± 23	49 ± 27	43 ± 25
TOUCH	[μm]	0.24 ± 0.12	0.37 ± 0.15	0.60 ± 0.30
DARKN	—	7.61 ± 1.52	6.49 ± 1.50	5.27 ± 1.33

[a] Parameter names refer to Table I; a.u. = arbitrary units.

TABLE III

Parameters with a Different Mean Value for the Classes Indicated at a Level of Significance of 1%

Negative *vs* dysplasia	: MEACON,[a] DIFF, DARK, DARKN, TOUCH
Negative *vs* invasive carcinoma	: MEACON, DIFF, DARK, DARKN, TOUCH
Dysplasia *vs* invasive carcinoma	: MEACON, DIFF, DARK, DARKN, TOUCH

[a] Parameter names refer to Table I.

TABLE IV

Preliminary Results of a Classification Trial on 112 Cervical Cell Nuclei from Three Different Cytologic Specimen Classes Using Texture Parameters

IVA. Testset is not equal to trainingset

a posteriori Class	*a priori* Class		
	Negative	Dysplasia	Invasive Carcinoma
Negative	17	2	3
Dysplasia	5	40	5
Invasive carcinoma	1	7	32

IVB. Testset is trainingset

a posteriori Class	*a priori* Class		
	Negative	Dysplasia	Invasive Carcinoma
Negative	19	2	3
Dysplasia	4	42	6
Invasive carcinoma	0	5	31

in the chromatin pattern it yields digitized images of 25k points or more requiring large computer capacity. In order to fit these images into the PDP 11/60 the data are overlayed by the TIC program. For a real time application this is not computationally efficient. Therefore it is essential to investigate whether a lower spatial or gray level resolution will give acceptable results.

The development and testing of texture analysis algorithms require a reliable cytologic *a priori* classification which is difficult to achieve for single cells. If a certain percentage of cells is not correctly classified *a priori* then the validity of the result obtained is influenced to the same degree. Notwithstanding this problem several investigators have reported promising results from analysis of cervical cells with various techniques (1, 7) although the extent to which these might have been influenced by the ambiguity of the data set is not completely clear. An attempt has been made to reduce this ambiguity by avoiding cytologic classification of single cells. An objective cell selection criterion (DNA content) is used instead and the specimen diagnosis for the *a priori* class for all cells. The use of a preselection criterion is based on the assumption that in a real-time system the vast majority of cells can be rapidly excluded from further analysis by one or more fast preselection criteria. In the Leyden television analysis system for automated cytology (LEYTAS) (12) preselection will be done on, *e.g.,* DNA content. The data reduction thus achieved will permit the more elaborate and time consuming analysis of the chromatin pattern and may lead to the final diagnosis of the specimen.

The segmentation algorithm described here has certain advantages over the use of only a fixed gray level threshold, or a threshold calculated for each nucleus individually, for the segmentation of the chromatin pattern (5). The first method requires a highly reproducible staining of the cells whereas the second method often leads to a loss of detail by merging of the different regions above the threshold. In our approach these regions with relative high chromatin contrast are found by a gradient procedure which is restricted to 20% of the nuclear image points by an O.D. threshold calculated for each nucleus individually. By using a fixed percentage of the nuclear area the effect of variations in staining intensity on the partitioning of the nuclear image is minimized.

Although empirically the chosen value of 20% gives a quite satisfactory correspondence between the segmented regions and the most conspicuous details in the chromatin pattern, the influence of the height of the threshold on the discriminatory power of the texture parameters derived from the segments needs further investigation.

Recently Yasnoff *et al.* (13) described two objective error measures for the quantitative evaluation of segmentation methods by comparison of the segmented image with a "true" or reference segmentation obtained by optical overlay of the original image on a representation of the segmented image. The performance of the objective error measures on segmentation in white blood cell analysis showed a good correlation with the human assessment of segmentation errors. Application of these objective error measures for optimization of the segmentation algorithm for cervical cells is hampered by the problem that the "true" segmentation of the regions with high chromatin density is less circumscript than, *e.g.,* that of nuclei in white blood cells.

The preliminary results obtained with the segmentation algorithm on the limited data set so far indicate that texture parameters can be derived that show quantitative differences in visually observable properties such as intensity and uniformity of staining, and spacing of condensed chromatin regions between nuclei from different cytologic specimen classes. The results from the classification trial indicate within the limitations set by the size and the composition of the data set that the six texture parameters tested permit the distinction between three groups of texture types with increasing abnormality. It is not surprising that nuclei from the invasive carcinoma specimens are allocated to the two other classes since in such specimens the presence of cells with less severe changes in chromatin pattern are not excluded. Also the selection of the nuclei was not particularly directed to include only the cells with the texture types most typical for invasive carcinoma as described by Patten (6) " . . . , irregularity of the distribution of the chromatin patterns." More data are needed to draw valid conclusions from the confusion between the *a*

priori dysplasia class and the *a posteriori* invasive carcinoma class which might be caused by a classification error or may point to the presence of cells with highly abnormal chromatin patterns in dysplasias. A possible approach to this problem could be the inclusion of the visually most abnormal patterns in each of the two classes. The same selection procedure should be used to investigate whether cells from carcinoma *in situ* specimens can be discriminated from cells derived from dysplasia and invasive carcinoma on the basis of textural differences. By use of the 5 C criterion for selection of the cells from carcinoma *in situ* specimens only we would expect to have too much overlap with the other classes. In addition the negative *a priori* class should be supplemented with more images from nuclei with a DNA content over 5 C occasionally found in negative specimens as shown in the work of Ploem-Zaaier *et al.*[1] since diploid nuclei of normal intermediate cells would not be selected in the ultimate real time system. With such a system the subjective element in the selection of the cells may be eliminated. For this reason a first version of TIC is now being implemented in the LEYTAS system for automated cervical cytology. In conclusion these preliminary results indicate that texture analysis by image segmentation has interesting potentials for objective texture description and for the automated classification of specimens in addition to the texture parameters currently in use.

ACKNOWLEDGMENTS

We would like to thank Mrs. M. Pilouw-Mulder for her skillful handling of the films, Dr. A. M. Vossepoel for critically reviewing the manuscript and Dr. J. P. Mitchell for correcting the English text.

LITERATURE CITED

1. Abmayer W, Borst H, Burger H, Schwarzkopf G, Bloss WH, Erhardt R, Reinhardt ER, Hermann G, Poppl-Otto S, Soost HJ: High resolution cell scanning and feature extraction in the TUDAB project, Poster Presentations at the 5th Eng. Found. Conf. on Automated Cytology. (ed. Sullivan). Pensacola, 1976, 50
2. Cornelisse CJ, Ploem JS: A new type of two color fluorescence staining for cytology specimens, J Histochem Cytochem 24:72, 1976
3. Gelsema ES: ISPAHAN, an interactive system for statistical pattern recognition, Proc Biosigma, Paris, 1978, 469
4. Groen FCA: Analysis of DNA Based Measurement Methods Applied to Human Chromosome Classification, Ph.D. Thesis, Delft, 1977
5. Kiefer G, Kiefer R, Moore GW, Salm R, Sandritter W: Nuclear images of cells in different functional states, J Histochem Cytochem 22:569, 1974
6. Patten SF: Monographs in clinical cytology, Volume 3, Diagnostic Cytopathology of the Uterine Cervix, 1978, pp 251
7. Poulsen RS, Oliver LH, Cahn RL, Louis C, Toussaint S: High resolution analysis of cervical cells—A progress report, J Histochem Cytochem 25:689, 1977
8. Pressman NJ: Markovian analysis of cervical cell images, J Histochem Cytochem 24:138, 1976
9. Tanke HJ, Van Ingen EM, Ploem JS: Acriflavine Feulgen stilbene staining, a procedure for automated cervical cytology with a television based system (Leytas), J Histochem Cytochem 27:84, 1979
10. Van der Ploeg M, Van den Broek K, Smeulders AWM, Vossepoel AM, Van Duijn P: HIDACSYS, computer programs for interactive scanning cytophotometry, Histochemie 54:273, 1977
11. Van der Ploeg M, Van Duijn P, Ploem JS: High resolution scanning densitometry of photographic negatives of human metaphase chromosomes I instrumentation, Histochemie 42:9, 1974
12. Vrolijk J, Pearson PL, Ploem JS: LEYTAS, a real time image analysis system for automated cytology, J Histochem Cytochem (in preparation)
13. Yasnoff WA, Mui JK, Bacus JW: Error measures for scene segmentation, Pattern Recogn. 9:217, 1977