INVARIANCE OF TEXTURAL FEATURES IN IMAGE CYTO- AND HISTOMETRY UNDER VARIATION OF SIZE AND PIXEL MAGNITUDE

Karsten Rodenacker

GSF Forschungszentrum für Umwelt und Gesundheit GmbH, Neuherberg
Institut für Pathologie; Ingolstädter Landstr. 1, D-85764 Oberschleißheim

ABSTRACT: Textural features of the granular structure of stained cell nuclei and nuclear sections derived from co-occurrence and run length matrices are often used for correlation with external (clinical) parameters. The representation of cell nuclei and nuclear sections vary considerably under changes of preparation and fixation conditions. Most obvious are changes in size e.g. by fixation as well as changes in the amount of bound stain. Computer simulated variations in size and pixel magnitude of a set of images of cell nuclei stained with Feulgen were featured and compared.

Keywords: Image analysis, pattern recognition, cytometry, histometry, texture analysis, co-occurrence features, run length features, feature invariance

INTRODUCTION

In image cyto- and histometry quantitative features of granular structures or textures respectively of cell nuclei are measured for correlation with cell and/or tissue properties. An often used set of methods is derived from considering textures in images as generated by a stochastical process. The estimated parameters of the probability density function of such a (2^{nd} order) process are quantitative textural features. Most well-known are the co-occurrence and the run length matrix as estimates of the probability density function, introduced by Haralick et al. [4] and Galloway [3]. A recent review with abundant reference list concerning these features can be found in [6].

Invariance of quantitative features from different influences and distortions in digital image analysis is one of the most important property of any analysis and forthcoming interpretation. Well-known influences in cyto- and histometry from specimen preparation and staining as well as from the material itself and from the procedures applied result in geometric and densitometric variations. Often such variations are of diagnostic value and reflect specific properties of the material under examination. However feature extraction methods might not react properly to such variations or distortions.

Feature extraction methods in digital image analysis have reached such a degree of complexity [7], that the analysis of the algorithms alone is not sufficient to find out reliable answers to the question of feature invariance [2]. In this paper the methods of feature extraction for co-occurrence and run length features are shortly stated, analyzed and applied to a set of digitized images of stained cell nuclei. Each original digital image from a stained cell nucleus is altered in terms of size and grey value and analyzed with the whole feature set. Alterations are performed by calculation in the digitized image. Feature values are plotted against the varied parameters. Classifications as in [5] were not performed. The goal was to find out how far experimental influences in a limited range on cell images can pretend differences in textural features. The algorithms applied are used in routine cell image analysis at our laboratory [7].

MATERIAL AND METHODS

In image cyto- and histometry the pixels in digitized images represent mostly a measure

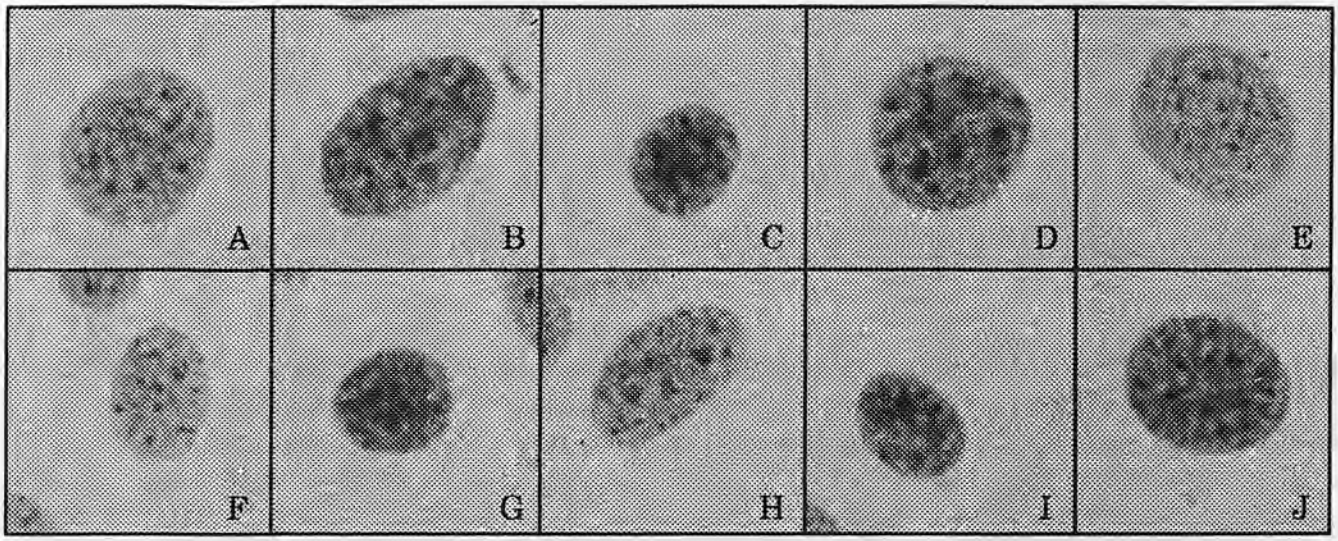


Fig. 1: Original digitized images of cell nuclei in transmitted light

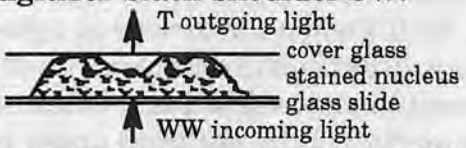
of optical density or extinction directly related to the physical amount of material (stain) at the location of the pixel. In this case after defining the nuclei of cells as regions of interest, the textural measurements are performed.

a – Image sampling and variation

From a larger project ten osteoblast cell nuclei (OBJECT=A,..,J) were arbitrarily chosen. The nuclei were stained according to our standard Feulgen procedure. Feulgen stain is deposited only at the DNA inside the cell nucleus, therefore the cytoplasm is transparent and mostly invisible. The amount of total stain in one nucleus is directly related to the amount of DNA, hence stoichiometric. The nuclei were digitized with an Axiomat microscope (Zeiss, Oberkochen, FRG) and a high resolution TV measurement camera (Bosch, Stuttgart, FRG, Plumbicon, T1VK9B1). Pixel size was 0.25 µm, picture size 128 × 128 pixel and the nominal resolution of grey values was 256. The selected original nuclei are shown in Fig. 1. Instead of extinction the transmitted light was measured, digitized and stored in an image file.

The DNA content is measured by the total extinction or integrated optical density respectively. Extinction is defined as the logarithm of the ratio of the transmitted outgoing light T and the incoming light WW, reduced by a digitizer black shoulder SW:

$$E = -150 \log_{10} \frac{T - SW}{WW - SW},$$



Since the amount and the distribution of DNA is of greatest interest in cyto- and histometry alterations of grey values were carried out in extinction. Pixelwise each calculated extinction value of each digitized image P_i, i=A,..,J (OBJECT), was multiplied by the factors β = {0.5, 0.6, 0.7, 0.8, 1.2, 1.4, 1.6, 1.8, 2.0} (FACTOR), recalculated into transmission values and than stored in an image file. The calculation was performed as a table calculation of exponential law according the formula without taking into account the black shoulder SW:

$$E' = -\log\frac{T'}{WW} = -\beta \log\frac{T}{WW} = \beta E \Rightarrow T' = WW^{(1-\beta)}\, T^{\beta}.$$

Subsequently this set of of 10x10 images $P_{i,\beta}$ were altered by size using an affine transformation with linear interpolation [9] resulting in 10x10x5 images $P_{i,\beta,\alpha}$. The x and y scaling factors were α = {0.5, 0.7, 1.0, 1.4, 2.0} (SIZE). Some dilated nuclei did not fit into 128 × 128 pixel image size and were truncated.

b – Feature extraction

Each of the cell pictures $P_{i,\beta,\alpha}$ were evaluated with a feature extraction procedure. First-

ly each image is transformed per pixel into normalized extinction according the above mentioned formula. Parameters controlling the feature extraction from the co-occurrence matrix are the displacement vector d (DISP), changed according the size factor β (SIZE), the type of normalization (NORM) of the grey values and the matrix size (NOGV). The probability marix of co-occurrences and run length's was calculated using the displacement vector size DISP on the grey value reduced normalized extinction image (with NOGV and NORM) in two directions to smooth out sampling artefacts (fig. 2). The feature numbering sceem follows Haral-

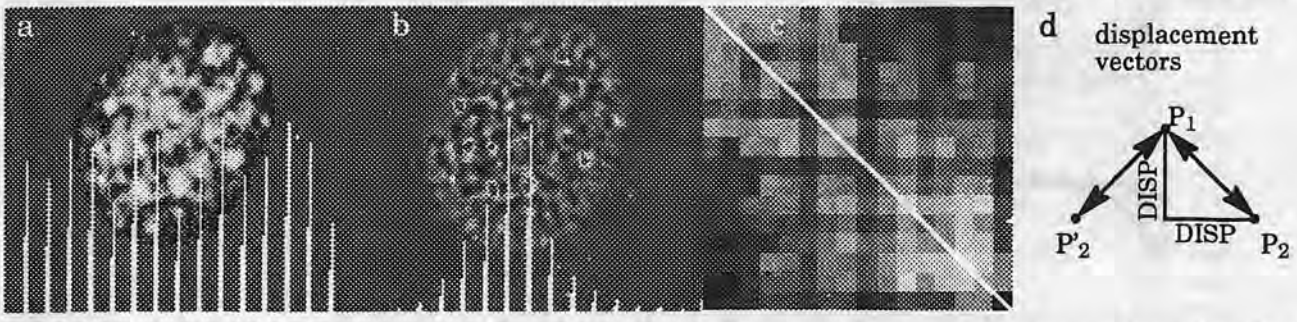


Fig. 2: Normalization with histogram equalization (original extinction a) and linear transformation (flat texture image b) with NOGV=16, probability matrix from a (c) and displacement vectors used (d)

ick [4] and Galloway [3]. Hence from each image $P_{i,\beta,\alpha}$ the co-occurrence features CO1-CO14 and NC1-NC14 were calculated for DISP=1, 3, 6, 10, 15; NORM=1 (linear), 2 (histogram equalization); NOGV= 8, 16, 32, 64. For run length features RL1-RL5 only the normalization NORM and matrix size NOGV was varied. Additionally the same algorithms were applied to a transformed image called *flat texture image* which is calculated as the difference of the (original) extinction image and the median filtered one. These features are abbreviated to NC1-NC14 and NR1-NR5 respectively.

c – Feature analysis

All 20000 sets of CO1-CO14 features and 4000 sets of RL1-RL5 features from the cell images were analyzed and displayed with the statistical analysis system SAS [8]. Most emphasis was given to illustrate invariance properties by variation of size and pixel magnitude.

RESULTS

a – Variation of pixel magnitude (FACTOR)

For SIZE=1.0, NORM=2, NOGV=16 and DISP=6 features CO1-CO14 and for SIZE=1.0, NORM=2 and NOGV=16 features RL1-RL5 are plotted against the magnitude varying parameter FACTOR (fig. 3). In spite of the rigid normalization several features show some dependency from pixel magnitude (FACTOR), see CO1, CO4, CO9 and RL2-RL5. Other apparently dependent features have very small ranges. These dependencies can be differentiated further on for groups of *bright* images (OBJECT=A, E, F, H), *medium* images (B, D, J) and *dark* images (C, G, I) (fig. 1). This is especially valid for CO1, CO8, CO9 where features from *bright* images decrease or increase more for FACTOR<1.0.

To illustrate the variation of features under changes of magnitude each feature was normalized by division by the corresponding feature calculated from the original image (FACTOR=1.0). In Fig. 4 with dotted lines the mean with ± 2 s.e.m. bars of the 10 objects for NORM=2 is shown. Reference lines for value "1" (=100%) are added. Small variations ($<\pm 10\%$) are recognizable for CO2, CO5-CO11, CO14, RL1, RL4, RL5 [vertical axis range 0.86 – 1.11). Medium variations ($<\pm 50\%$) appear for CO3-CO4, CO12, RL2, RL3 and large variations (about $\pm 100\%$) only for CO1.

Finally each feature should define an order relation above the set of images. The invari-

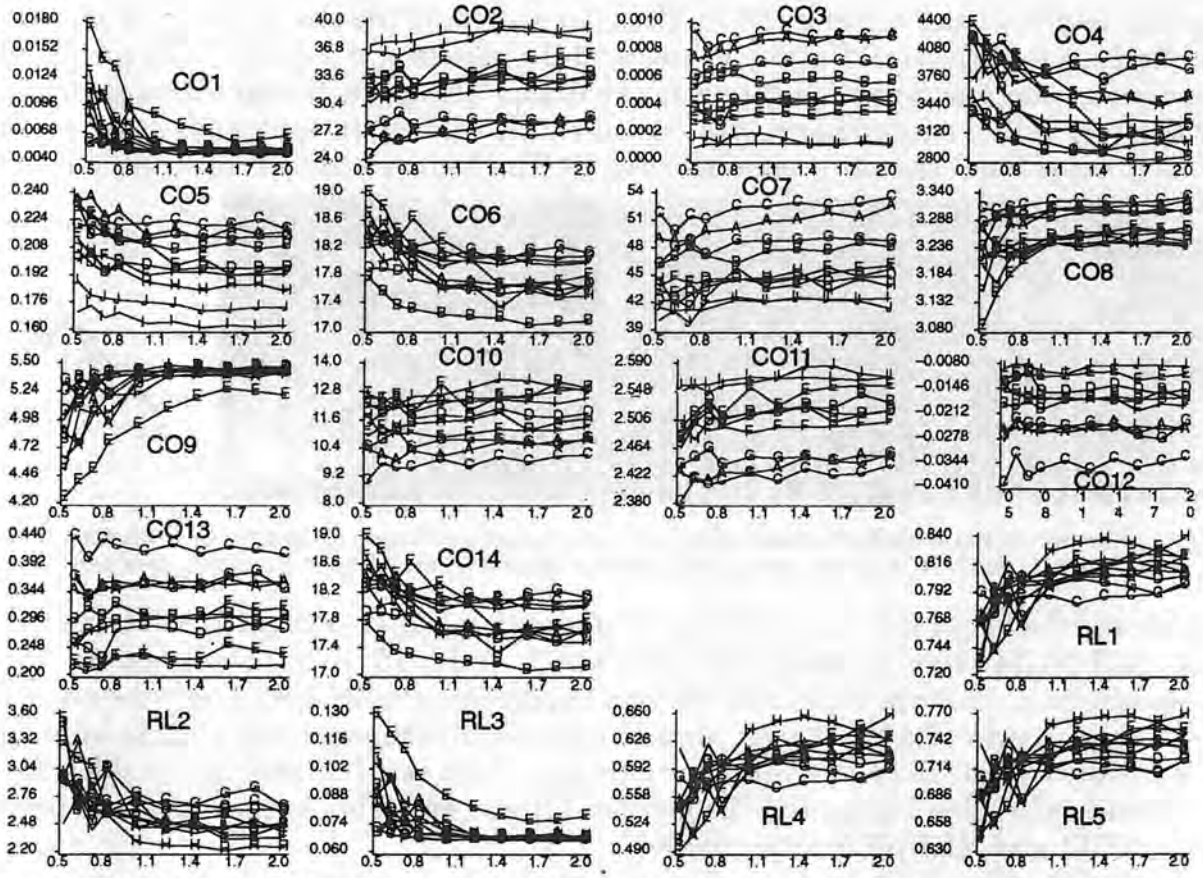


Fig. 3: All CO-features for SIZE=1.0, NOGV=16, DISP=6, NORM=2 and all RL-features for NOGV=16, DISP=6, NORM=2 above FACTOR

ance of the latter holds for CO2, CO3, CO5, CO7, CO8, CO10-CO13 with some individual exceptions. This can easily be recognized by the small number of crossing lines in the corresponding plots (fig. 3). The features of different parameters of SIZE, DISP, NORM and NOGV show similar behavior.

b – Variation of pixel size (SIZE)

The same procedure as stated in the previous paragraph was applied to show the influence of scale changes. For FACTOR=1.0, NORM=2, NOGV=16 and DISP=6 features CO1-CO14 and for FACTOR=1.0, NORM=2 and NOGV=16 features RL1-RL5 are plotted against the size varying parameter SIZE (fig. 5). The dependencies are in this case much larger compared to the previous ones. Largest influences exist for RL1, RL2, RL4, RL5 and CO2, CO3. The dependency of RL-features is expected since run length is directly related to size. The strange behavior of CO7 and CO10 seems to be a result of the image interpolation during scale changes interfering with the integer adjustment of displacement vector size (DISP*SIZE).

To illustrate also the variation of features under changes of size each feature was divided by the corresponding feature calculated from the original image (SIZE=1.0) again. In Fig. 6 mean and ± 2 s.e.m. bars of the 10 objects are shown. Large variations ($\pm 100\%$) appear for CO3, CO7, CO10, CO12. Medium variations ($<\pm 50\%$) exist in CO1, CO2, CO4, CO5, CO13 and small variations appear in CO6, CO8, CO9, CO11, CO14.

No order relationship could be recognized in the set of images in Fig. 5.

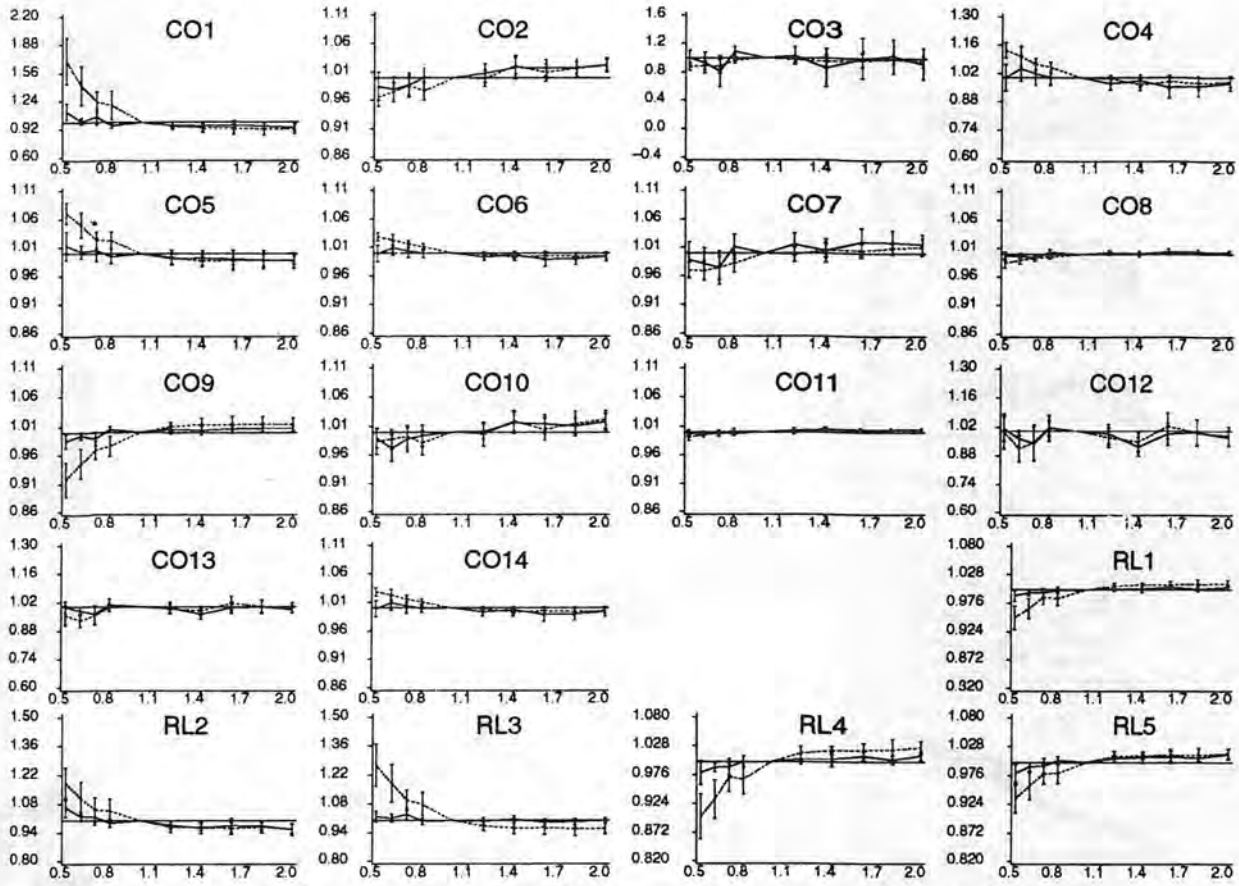


Fig. 4: Mean and ± 2 s.e.m. from all objects for NORM=1 (solid), 2 (dotted) for all CO-features for SIZE=1.0, NOGV=16, DISP=6 and all RL-features for SIZE=1.0, NOGV=16 normalized by the corresponding feature from FACTOR=1.0 plotted above FACTOR

c – Variation of grey value normalization (NORM)

In Fig. 4 and Fig. 6 all normalized features are plotted above FACTOR and SIZE for NORM=1 (solid) and NORM=2 (dotted). For orientation reference bars for *normal* (=1.0) are added. The variation of the features for the two methods of grey value normalization are similar. For very small ranges of grey values (FACTOR<1.0) the histogram equalization (NORM=2) is less efficient. More deviations from "1", the invariance, with larger magnitude appear. From Fig. 3 and Fig. 4 a larger dependency of NORM=2 features from FACTOR than for NORM=1 features is recognizable. However an order relationship is for CO- and RL-features less preserved (without additional Fig.). This is in contrast to the NC-features, see paragraph f).

d – Variation of matrix size (NOGV)

In general the variation of matrix size NOGV do not change the feature behavior. The features CO1, CO5, CO6, CO8, CO11, RL1 change their range of values but not their distribution. The features CO2, CO4, CO7, CO9, CO10, CO12, CO13, RL3-RL5 spread out with increasing matrix size. Conversely, the features CO3 and RL2 collapse. Features CO12 and CO13 are sensitive for the size of the region of interest. The outlying OBJECT=C, I, G are the smallest nuclei and OBJECT=E is the largest one. For illustration the features CO1, CO2, CO3 and CO13 are shown in Fig. 7.

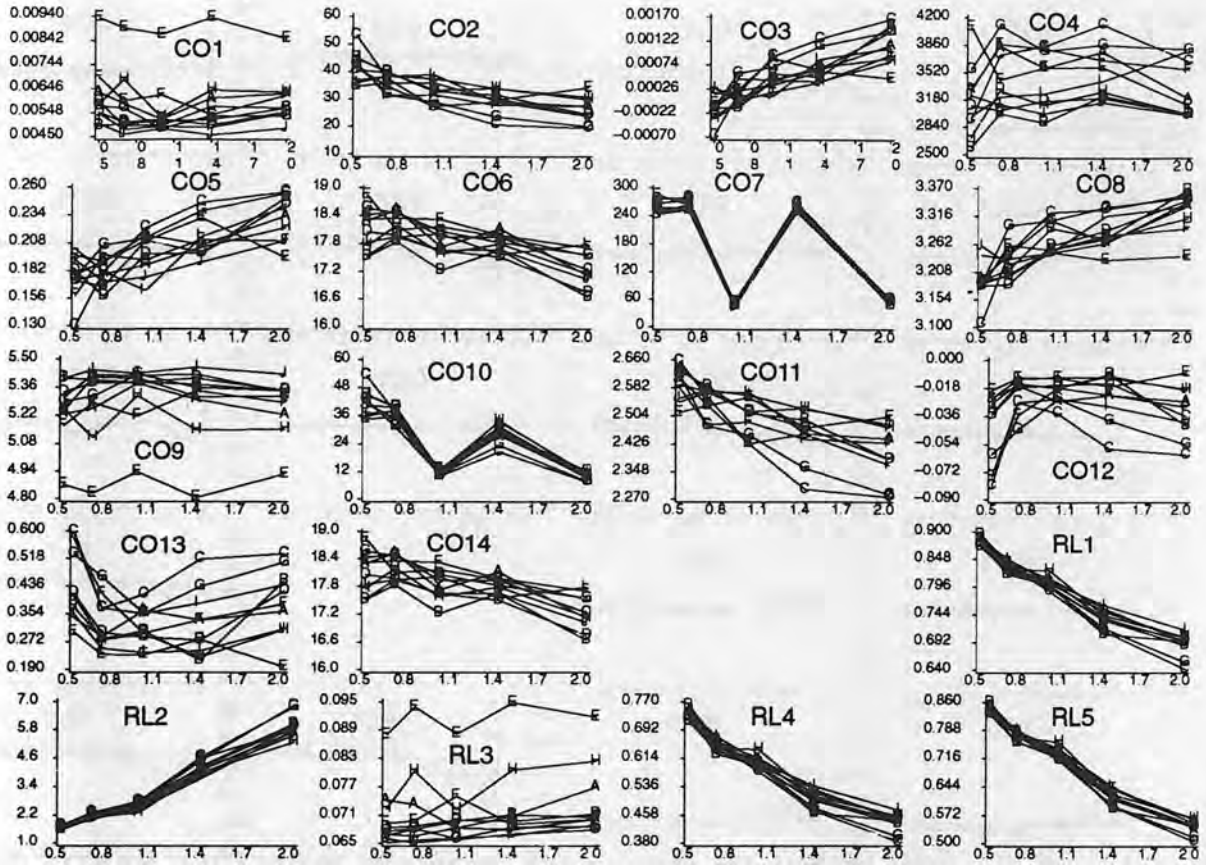


Fig. 5: All CO-features for FACTOR=1.0, NOGV=16, DISP=6, NORM=2 and all RL-features for FACTOR=1.0, NOGV=16, NORM=2 plotted above SIZE

e – Variation of displacement vector size (DISP)

The variation of the size (and/or of the direction for non-isotropic textures) of the displacement vector is usually described as resulting in a variation of sensitivity of CO-features for certain textures. In our experiment the number of different textures is quite limited and by no means distinct as in [5] or in the often used set of example images from [1]. However the graphic representation of the features as a function of DISP show that at least CO1 and CO9 do not reflect any different sensitivity (similar behavior for each object, order relationship quite stable, nearly no line crossings). All other features show only minor differences for DISP$\leq$6 and DISP$\geq$10. This is even less pronounced for features calculated with NORM=1. Small deviations appear in nearly all CO-features for OBJECT=C, F, G, I. These are just the smallest objects. For illustration the features CO1, CO3, CO4 and CO9 are shown in Fig. 8.

f – Features derived from flat texture image (NC-features)

For features derived from a transformed image (*flat texture image*) no figures are displayed. They show comparable development, except for the two different methods of grey value normalization (NORM). For the NC-features the linear transformation (NORM=1) is more efficient. This is valid in terms of dependency from FACTOR and in preservation of order relationships (NC3, NC7, NC9, NC13). The order relationships differ from those defined by CO-features. In terms of variation of the displacement vector (DISP) only sizes less or equal the window size of the median transformation are adequate. The median window size was not altered for different SIZE parameters in this experiment.

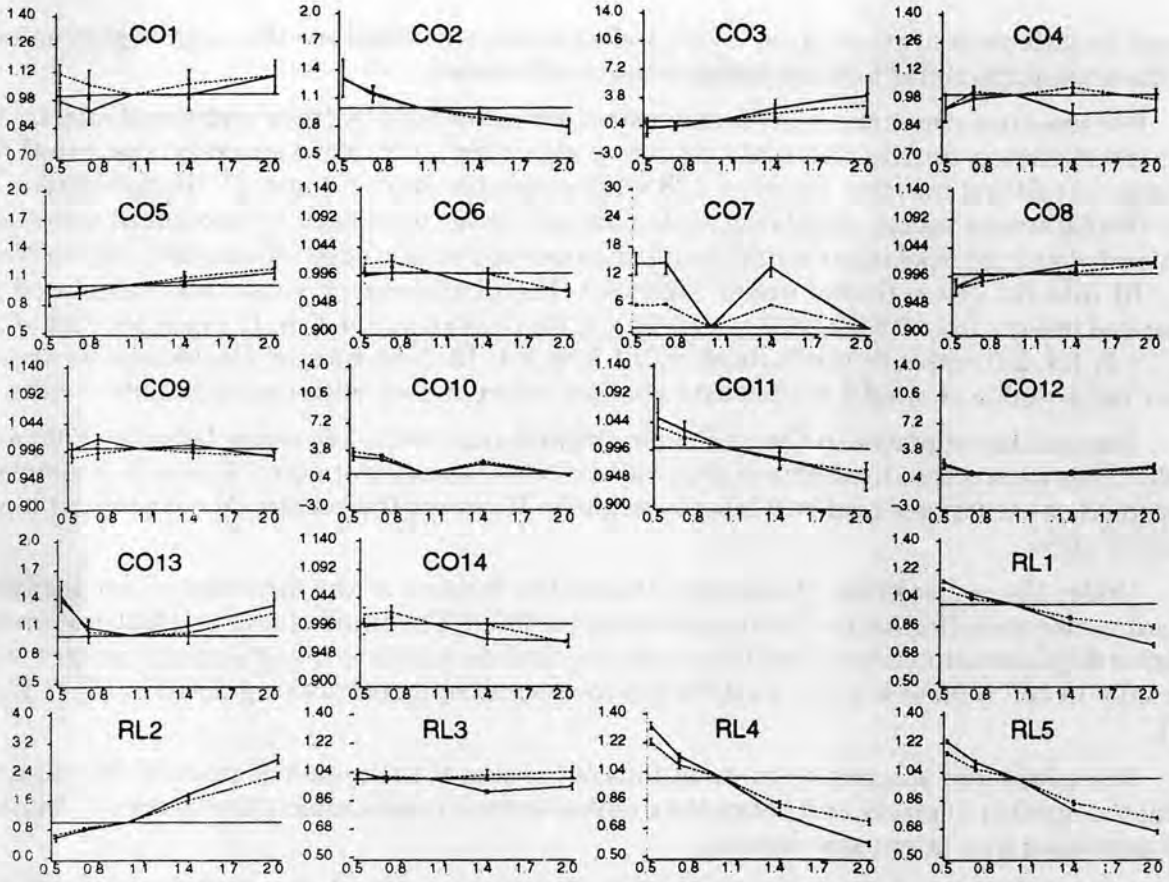


Fig. 6: Mean and ± 2 s.e.m. bars from all objects for all CO-features for FACTOR=1.0, NOGV=16, DISP=6 and all RL-features for FACTOR=1.0, NOGV=16 normalized by the corresponding feature from SIZE=1.0 plotted above SIZE for NORM=1 (solid), 2 (dotted).

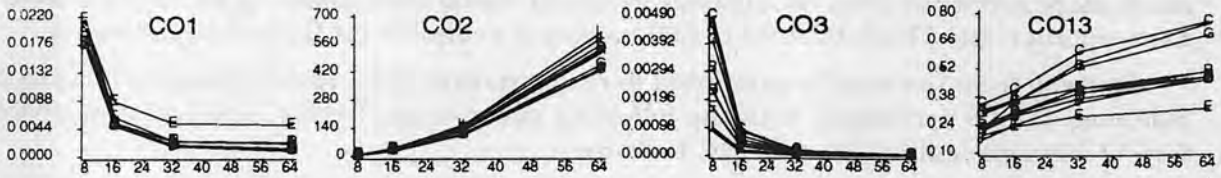


Fig. 7: Selected CO-features for FACTOR=1.0, SIZE=1.0, DISP=6, NORM=2 plotted above NOGV

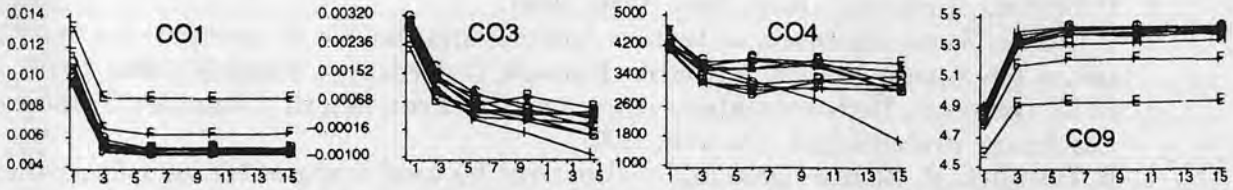


Fig. 8: Selected CO-features for FACTOR=1.0, SIZE=1.0, NOGV=16, NORM=2 plotted above DISP

DISCUSSION AND SUMMARY

From all varied parameters only some results are displayed. The most surprising properties are the frequently existing dependencies of features from grey value variations for images with low grey values (FACTOR<1.0). We suppose that the process of normalization by equalization is not efficient enough for small ranges of grey values. With other words: if there is only a small number of different grey values, the equalization will and cannot perform properly. This would explain the better performance of linearly normalized NC-fe-

tures. As difference of an original image and a median smoothed one the range of grey values in the transformed flat texture image is naturally small.

For size alterations the resulting features are influenced by three additional effects: 1st the interpolation during the scale changing algorithm, 2nd (more severely) the cutoff for images not fitting into the frame of 128×128 pixels for SIZE>1.0 and 3rd the calculation of the displacement vector. Applying scale changes using minimum or maximum operators instead of any interpolation would result in unwished strong edge effects. Only OBJECT=C, G, I fit into the given frame under SIZE=2.0. The displacement vector was calculated as rounded integer from DISP*SIZE resulting e. g. for DISP=6 in 3, 4, 6, 8, 12 pixels for SIZE=0.5, 0.7, 1.0, 1.4, 2.0 respectively instead of 3, 4.2, 6, 8.4, 12. Non-integer displacements are either not possible or would necessitate another interpolation with unpredictable results.

The number of pixels in the nuclei for original size (SIZE=1.0) range between 1700 and 4200. This means that numbers of grey values (NOGV) greater or equal 32 result in sparsely occupied co-occurrence and run length matrices. However the results do not prohibit large matrix sizes.

Under the assumption of isotropic objects the fixation of the direction of the displacement vector and alteration of its magnitude is justified. The small changes of features under varied displacement vector sizes let us suppose that the textures in our example images (and usually in cell nuclei) are not as different as macroscopic textures e.g. used in [5] or from [1].

Not taken into account were variations of the size of the region of interest. We suppose that the number of pixels and hence the number of occurrences affects the features. This will be examined in a later experiment.

At least for larger grey values most of the features examined show satisfying invariance from changes in grey value. Also the order relationship is relatively stable for the chosen texture examples as far as they can be representative. For size changes the features are much more scattered. May be a change of size in continuous space e.g. by optical means is more appropriate. The actual set of 500 images is available for further experiments.

Derived from the results presented here our routine cyto- and histometry feature calculations will be performed with the following parameters: DISP=6, NOGV=16 and NORM=2 for CO-features and NORM=1 for NC-features.

REFERENCES

1. P. Brodatz, Textures, Dover, New York, 1966.
2. P. Carter, Some comments on texture analysis and the use of co-occurrence matrices, UK Atomic Energy Authority Harwell, Oxfordshire, **AERE M-3043**, 1979.
3. M.M. Galloway, Texture analysis using gray level run length, Computer Graphics and Image Processing **4**, 172-179, 1975.
4. R. Haralick, K. Shanmugam and I. Dinstein, Textural features for image classification, IEEE Trans. Syst. Man Cybern. **SMC-3**, 610-621, 1973.
5. P.P. Ohanian and R.C. Dubes, Performance evaluation for four classes of textural features, Pattern Recognition **25**, 819-833, 1992.
6. T.R. Reed and J.M. Hans du Buf, A review of recent texture segmentation and feature extraction techniques, CVGIP: Image Understanding **57**, 359-372, 1993.
7. K. Rodenacker, M. Aubele, G. Burger, P. Gais, U. Jütting, W. Gössner, M. Oberholzer, Image cytometry in histological sections of colon carcinoma, Acta Stereologica **11/Suppl 1**, 249–254, 1992.
8. SAS-Manuals SAS/STAT, SAS/GRAPH, SAS Institute Inc. Cary, 1990.
9. SPIDER User's Manual, JSD, Tokyo, 1983.

ICAPRDT'93

28—31 December, 1993
Calcutta INDIA

Indian Statistical Institute
203 B. T. Road, Calcutta 700 035

PROGRAMME

P C Mahalanobis
Birth Centenary Volume

onsors:
ernational Association for Pattern Recognition, Inc.