

REVIEW ARTICLE

Image analysis and definition of nuclear phenotypes*

Benedicto de Campos Vidal

ABSTRACT

Some basic topics concerned with the extraction of textural and geometric information from cell nucleus images as well as description and characterization of chromatin supraorganization and consequent classification of nuclear phenotypes are presented.

INTRODUCTION

Image analysis is used to extract information contained in images. Image analysis aims at detecting, diagnosing and describing the texture and geometry of the images. The information content and, consequently, the meaning of an image are function of the processes used for its generation and acquisition. The features through which an image is identified and categorized are of spectral and geometric (or morphometric) nature.

In the case of cell nuclei and chromatin supraorganization, image analysis aims at obtaining data concerned with textural and geometric characteristics that will define nuclei as individuals pertaining to a precise class with specific physiological meaning.

An important historical fact is the report made by Jacoby in 1925 and 1935 that an increase in nuclear size corresponds statistically to classes with a growth rhythm obeying the 1:2:4:8:16 rate, which would be linked to cell physiology (Bucher and Horisberg, 1950). In this case and in subsequent investigations by other authors morphometric methodology was used. Since then, the methods for the study of nuclear and nucleolar geometric characteristics have evolved. Modern

morphometric procedures involve the use of digitising tables or TV-video image analysis systems.

METHODS

Nuclear images may be obtained in several ways, each of them promoting the acquisition of different types and amounts of information. Fresh unfixed material may be analyzed by phase contrast microscopy. Interference microscopy furnishes data on dry masses of nuclear material (DNA and proteins). Cytochemical or immunocytochemical methods produce images that diagnose types of chemical substances that may be of interest.

Video image analysis comprises various steps, as follows: Image capture, made by a CCD camera that produces analogical signals of voltage that are proportional and correspondent to brightness and grey levels in the different parts of the object (cell nuclei, in the present case). These signals are converted into digital (that is, numerical) information by a processor (frame grabber). The numerical values corresponding to the various brightness or grey shades are positioned between zero (black) and 255 (white).

A pixel is a picture unit or element. It is the smallest area outlined in the scanned surface. It may be roughly compared with silver grains in a photograph.

* Conference presented at the 42nd National Congress of Genetics. September 4-7, 1996, Caxambu, MG, Brasil. Departamento de Biologia Celular, Instituto de Biologia, UNICAMP, 13083-970 Campinas, SP, Brasil.

The images may be captured from photographs, micrographs or directly from the microscope.

For images taken from the microscope, some basic requirements should be met. These are a stable light source, absence of voltage instability, homogeneous illumination, according to Koehler, and weak illumination in the microscope room in order to prevent reflectance phenomena. Even the microscope design should be such that undesirable optical phenomena are not produced.

Image treatment

This step is intended to produce an image with a good delimitation of its textural characteristics, that is, with a homogeneous distribution of grey levels. Visibility and other qualities of the image are a function of its spectral and geometric properties. Absorbance (also called optical density, OD), refraction and reflectance are spectral properties. Under these terms, the use of monochromatic filters is mandatory. Selection of a wavelength, λ , for which the structures under study will exhibit an absorption maximum or will appear close to the peak of the absorption curve, will configure a larger probability of acquisition of identifiable grey levels. In fact, since absorption is sometimes very intense, a λ far from the spectral region containing the absorption peak should be chosen in order to avoid measurement errors and apparent homogeneity of the object.

When the image is captured, the signals pass through the frame grabber. There is a table named LUT (= look-up table) or ILUT (= input LUT) which corresponds to the work area in the monitor. In this work area it is possible to get more or less image contrast and brightness and to save LUT data for standardizing all of a series of measurements. After being captured, the image may be operated through a display tool that will allow more or less contrast, or even artificial colors.

Profiles and histograms of the gray levels of the nuclear images

As an initial step for the evaluation of grey level variability and possible nonhomogenic distribution of grey levels it is recommended that profiles and histograms of the grey levels of the nuclear images are made. It is also useful for the establishment of some parameters for image segmentation. Figures 1 and 2 explain the information extent obtained in this phase. The histograms are accompanied by statistics of the grey levels. Emphasis should be given to standard deviation values (StDv). The data obtained from these

histograms are a first step for quantification of the variability of the average grey distribution, i.e. OD, allowing the classification of the chromatin suprastructure to be in agreement with the homogeneity or heterogeneity of the chromatin absorption rates in the nuclear topography. The mathematical descriptor preferentially used in this case is the standard deviation of the OD values, referring to all nucleus pixels. The StDv mean of the nuclear population has been named diffuseness (Einstein *et al.*, 1994).

The discrimination of the image, separating its structure from connections with other structural elements at the background is called segmentation. After the segmentation of one or several nuclei within a work field it may occur that a new segmentation of areas within the nuclei (heterochromatin or condensed chromatin, for instance) must be performed (Figures 3 and 4).

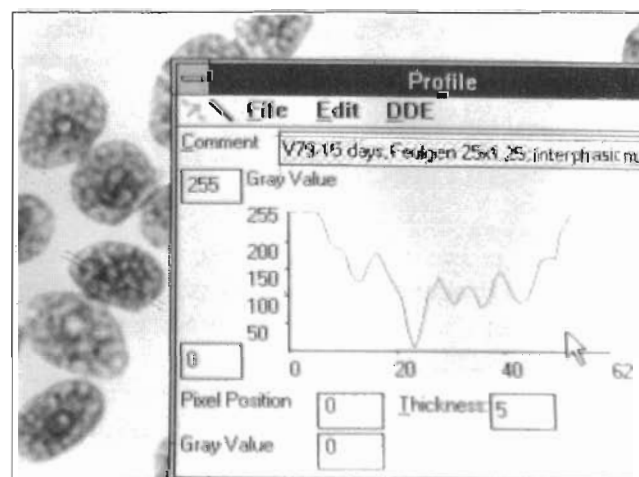


Figure 1 - Feulgen staining distribution profile in a V79 cell. Crossing the nucleus with a measuring area 5 pixels wide over 5 condensed chromatin spots corresponds to 5 troughs in the profile curve, the gray levels of which are considered in the abscissa. The more condensed spots correspond to 25.6 pixels. Each pixel equals 0.2244 μm (X axis).

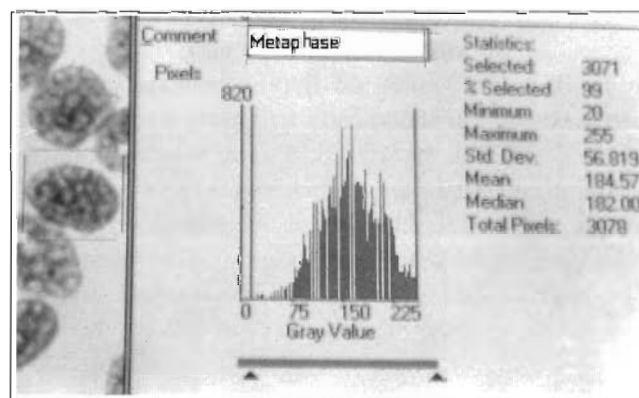


Figure 2 - Gray level histogram for a Feulgen-stained V79 cell nucleus, including some unstained surrounding area. The corresponding statistics are given at right. The standard deviation (Std.Dev.) corresponds to a coefficient of variation of 31%, to which values over 160 pixels (outside the nucleus) contribute.

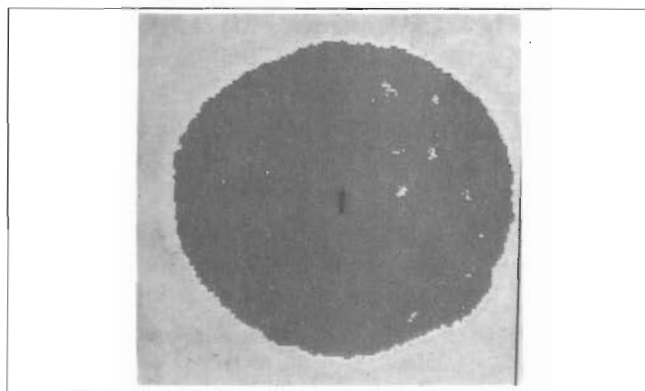


Figure 3 - Binary image of a Feulgen-stained liver cell nucleus after discrimination of the nuclear boundaries. The image was taken directly from the monitor.

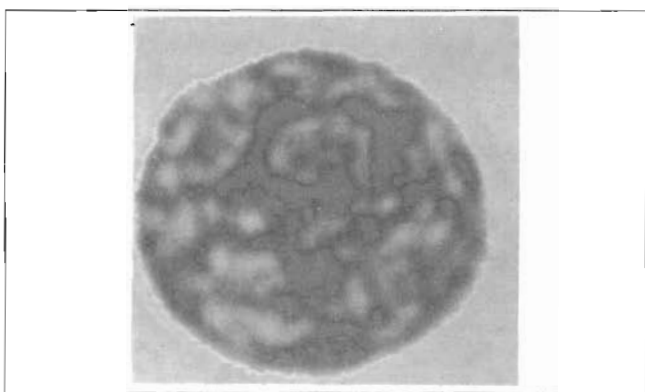


Figure 4 - Binary image of a Feulgen-stained liver cell nucleus after the clumps of condensed chromatin. The image was taken directly from the monitor.

Mathematical descriptors and pattern recognition

There are nearly 70 mathematical descriptors used for the recognition of nuclear patterns. Most of them are derived from basic descriptors extracted directly from image features. From a practical point of view, the use of a relatively small number of descriptors extracted directly from the image may be recommended. The use of nearly a hundred descriptors obtained from mathematical or arithmetical manipulations has been more a source of confusion than a clarifying key for solving cytometry questions.

Absorbance is the right term to designate the phenomenon which produces the grey levels of the image. However, the denomination "optical density" (OD) has been frequently more used. The image observed in the monitor display is a matrix of grey values. The segmented and digitized nucleus image (binary image) is thus expressed in total absorbance or total OD, resulting from the integration of the grey level values of the segmented area (IOD, integrated optical density). $OD (= IOD/area)$ is called average grey level. As a corollary, $IOD = OD \times area$.

The interpretation of the image texture will depend on the establishment of geometric descriptors. Among these, some may be cited:

- Area, arbitrarily denoted as S ;
- Perimeter: P ;
- Roundness, $(4\pi \times S)/P^2$, or perimeter form factor (Ff). For a circle, $Ff = 1$; for other shapes, $Ff < 1$.
- Circumference ratio (Vidal *et al.*, 1973), $P^2/(4\pi \times S)$, is the inverse of roundness and expresses how much an image shifts from a circle. The denomination "form factor" is not restricted to these two expressions. Some authors include the ratio smaller axis/larger axis to express a form factor especially designed for ellipsoidal shapes.

There are dozens of geometric descriptors directly or indirectly calculated and programmed to comply with various programs. Global Lab Image™ DT offers the possibility for calculation of 57 different geometric descriptors.

The OD standard deviation is a direct measurement of texture. Histograms of texture characteristics can be used for the detection of kurtosis and skewness.

I have proposed that a ratio similar to the variance, F , should be used such that the "F" ratio is a measure of the variability of the average absorbances. Consequently, $F = (StDv_{OD} \text{ among nuclei}) / (\text{Average } StDv \text{ for OD inside each nucleus})$.

Another proposition for characterization of nuclear texture and phenotypes (Vidal, 1984; Vidal *et al.*, 1984) is a scatter diagram relating $S_c\%$ (percentage of nuclear area covered with chromatin), displaying an OD value above a certain absorption level or cut off point, that is OD_{cf} to AAR (average absorption ratio, OD_{cf}/OD , Vidal *et al.*, 1973). It has proven to be a useful tool for comparison of nuclear phenotypes under several different experimental conditions (Mello, 1989; Mello and Russo, 1990; Vidal, 1992; Mello *et al.*, 1992, 1994, 1995; Mello and Chambers, 1993). An adequate statistical analysis of the obtained data is crucial for interpretation of the results.

RESULTS AND DISCUSSION

Some authors consider a variant of this technique as a Feulgen reaction, substituting thiazynic dyes for basic fuchsin (pseudo-Feulgen reaction). Under these circumstances the staining mechanisms may differ and results may also vary as a function of the tissue or organ under investigation. The variability in information contained in the image, when different cytochemical assays like Feulgen reaction, toluidine blue, gallocyanin and Giemsa are used, has been studied (Vidal *et al.*, 1973).

Standards have usually been employed for studies of quantitative variations in DNA amounts, especially when polyploidy is investigated. Diploid nuclei of the same organism have been recommended and considered as an internal control. Chicken erythrocyte nuclei (NEF) are often used as an external standard, which has also been the cause of some imprecision. Chicken erythrocyte nuclei are used as a standard for the calculation of dry mass (DM) of whole nuclei and DNA, for comparison of chromatin packing state levels. Altman and Katz (1976) have reported data on NEF from which I have calculated that the DNA DM is equal to $2.550 \text{ pg} \pm 0.425 \text{ pg}$, for $N = 11$. Image analysis data, obtained from the interference microscopy images, are in agreement with these data (DNA DM equal to $2.960 \text{ pg} \pm 0.276 \text{ pg}$, for $N = 151$). It was also possible to calculate, by image analysis, that there was a 46% loss of total nuclear DM as a consequence of Feulgen's acid hydrolysis. The chicken erythrocytes

Character stem-and-leaf display

Stem-and-leaf of μm^2 N = 180
Leaf unit = 1.0

A

```

1      2 8
7      3 337789
20     4 0155566677789
31     5 12233445677
54     6 01112223345556677788999
77     7 0000112344455555556668
87     8 0011112455
(8)    9 01678889
85    10 01144455566788
71    11 00012222333334444557788899999
41    12 0233345579
31    13 111357788999
19    14 00155
14    15 1
13    16 5588
9     17 18
7     18 9
6     19 4
5     20 37
3     21
3     22 0
2     23
2     24
2     25 2
1     26
1     27
1     28
1     29
1     30 1

```

Character histogram

Histogram of μm^2 N = 180

Midpoint	Count
25.0	1 *
35.0	6 *****
45.0	13 *****
55.0	11 *****
65.0	23 *****
75.0	23 *****
85.0	10 *****
95.0	8 *****
105.0	14 *****
115.0	30 *****
125.0	10 *****
135.0	12 *****
145.0	5 *****
155.0	1 *
165.0	4 ****
175.0	2 **
185.0	1 *
195.0	1 *
205.0	2 **
215.0	0
225.0	1 *
235.0	0
245.0	0
255.0	1 *
265.0	0
275.0	0
285.0	0
295.0	0
305.0	1 *

Figure 5 - Exploratory stem-and-leaf (A) and character histogram (B) statistics of the nuclear area in μm^2 . Basically, both histograms contain the same information. The polymodal distributions are due to the 2C, 4C and 8C classes determined in the integrated optical density (IOD) (Feulgen-DNA) frequency distributions (see next figures).

which enter the blood stream do not yet present nuclear pyknosis; their chromatin thus exhibits an OD variability that could be considered a model for the evaluation of chromatin texture heterogeneity (early apoptosis steps?). The standard deviation of the OD of the Feulgen-stained chicken erythrocytes was 1.258 (among nuclei) and 0.243 (within the nuclei, for data obtained from histogram measurements on each nucleus), for $N = 354$. Consequently, $F = 1.258/0.243 = 5.2$. This reflects the heterogeneity degree of chromatin texture, which was larger for comparisons among cell nuclei.

The basic procedures for nucleus image analysis can be demonstrated with another model, such as mouse or rat liver cell nuclei. These are considered good models in terms of their typical ploidy variability and chromatin supraorganization patterns (Vidal *et al.*, 1994; Smith *et al.*, 1996). Descriptive statistics like those provided by many statistical programs should be obtained. Since liver cell populations are polymodal, arithmetic means of areas and IOD are generally not representative of the nuclear populations. Examination of "stem & leaves" graphs as well as of character histograms gives information on the distribution and frequency of the data (Figures 5A,B and 6A,B). The same

Character stem-and-leaf display

Stem-and-leaf of IOD N = 180
Leaf unit = 1.0

A

```

1      0 4
2      0 5
4      1 03
33     1 56677777777888888899999999999
83     2 0000000000000111111111111111111222222222223344
85     2 56
86     3 0
(26)   3 556667777788888899999999999
68     4 0000000000000000011111111122222222222334444
24     4 5677777889
14     5 033
11     5 67
9      6
9      6 5
8      7 3
7      7 667
4      8 04
2      8 7
1      9
1      9
1     10
1     10 5

```

Character histogram

Histogram of IOD N = 180
Each * represents 2 obs.

Midpoint	Count
4.00	2 *
9.00	1 *
14.00	4 **
19.00	52 *****
24.00	25 *****
29.00	2 *
34.00	3 **
39.00	43 *****
44.00	26 *****
49.00	9 *****
54.00	3 **
59.00	1 *
64.00	1 *
69.00	0
74.00	1 *
79.00	4 **
84.00	1 *
89.00	1 *
94.00	0
99.00	0
104.00	1 *

Figure 6 - Stem-and-leaf (A) and character histogram (B) for Feulgen-DNA values in IOD units. In both histogram types polyploidy is clearly evident. IOD values in stem-and-leaf display are printed directly on the graphs. 2C values correspond to 14-25 IOD units, 4C to 30-50 IOD units, and so on.

- assortment of nuclear phenotypes based on chromatin texture evaluation. *Acta Histochem. Cytochem.* 27: 23-31.
- Mello, M.L.S., Content, S., Vidal, B.C., Planding, W. and Schenck, U.** (1995). Modulation of *ras* transformation affecting chromatin supraorganization as assessed by image analysis. *Exp. Cell Res.* 220: 374-382.
- Oberholzer, M., Oestreicher, M., Christen, H. and Bruehlmann, M.** (1996). Methods in quantitative image analysis. *Histochem. Cell Biol.* 105: 333-355.
- Smith, P.S., Parkinson, I.H. and Leong, A.S.-Y.** (1996). Principles of ploidy analysis by static cytometry. *J. Clin. Pathol.: Mol. Pathol.* 49: 104-111.
- Vidal, B.C.** (1984). Polyploidy and nuclear phenotypes in salivary glands of the rat. *Biol. Cell* 50: 137-146.
- Vidal, B.C.** (1992). Human renal carcinoma: Feulgen-DNA area and chromatin condensation determined by scanning cytometry. *Rev. Bras. Genét.* 15: 945-961.
- Vidal, B.C., Schlueter, G. and Moore, G.W.** (1973). Cell nucleus pattern recognition: influence of staining. *Acta Cytol.* 17: 510-521.
- Vidal, B.C., Silva, W.J. and Strikis, P.C.** (1984). Nuclear phenotypes and DNA content of root cells of *Zea mays mays*, *Zea diploperennis* and of a mazoid hybrid. *Cell. Mol. Biol.* 30: 11-22.
- Vidal, B.C., Planding, W., Mello, M.L.S. and Schenck, U.** (1994). Quantitative evaluation of AgNOR in liver cells by high-resolution image cytometry. *Anal. Cell. Pathol.* 7: 27-41.