Automatic morphological sieving: comparison between different methods, application to DNA ploidy measurements

Christophe Boudry [a,b,c,*], Paulette Herlin [b],
Benoit Plancoulaine [e], Eric Masson [d],
Abderrahim Elmoataz [d], Hubert Cardot [f],
Michel Coster [a], Daniel Bloyet [d]
and Jean-Louis Chermant [a]

[a] *Laboratoire d'Etudes et de Recherche sur les Matériaux, UPRESA CNRS 6004, Institut des Sciences de la Matière et du Rayonnement, Bld du Mal Juin, 14050 Caen Cedex, France*
[b] *Service d'Anatomie Pathologique, Centre Régional François Baclesse, Caen, France*
[c] *Laboratoire de Neurosciences, UMR CNRS 6551, Caen, France*
[d] *Groupement de Recherches et d'Etudes en Instrumentation, Informatique et Images, UPRESA CNRS 6072, Caen, France*
[e] *Institut Universitaire de Technologie, Département de Mesures Physiques, Caen, France*
[f] *Laboratoire Universitaire des Sciences Appliquées de Cherbourg, France*

The aim of the present study is to propose alternative automatic methods to time consuming interactive sorting of elements for DNA ploidy measurements. One archival brain tumour and two archival breast carcinoma were studied, corresponding to 7120 elements (3764 nuclei, 3356 debris and aggregates). Three automatic classification methods were tested to eliminate debris and aggregates from DNA ploidy measurements (mathematical morphology (MM), multiparametric analysis (MA) and neural network (NN)). Performances were evaluated by reference to interactive sorting. The results obtained for the three methods concerning the percentage of debris and aggregates automatically removed reach 63, 75 and 85% for MM, MA and NN methods, respectively, with false positive rates of 6, 21 and 25%. In-

formation about DNA ploidy abnormalities were globally preserved after automatic elimination of debris and aggregates by MM and MA methods as opposed to NN method, showing that automatic classification methods can offer alternatives to tedious interactive elimination of debris and aggregates, for DNA ploidy measurements of archival tumours.

Keywords: Automatic classification, DNA ploidy, image cytometry, mathematical morphology, multiparametric analysis, neural network

1. Introduction

One of the main problems encountered in image cytometry DNA ploidy measurement (ICM-DNA) of archival tumours is the purity of samples to be analysed. The preparation of dewaxed samples requires enzymatic dissociation of tissue, followed by the sedimentation of nuclei on slides. Although this technique is well standardised and reproducible [17], it generates much debris and aggregates (half of the total events on average). It has been previously shown that these debris and aggregates introduce a considerable bias on DNA ploidy measurements and consequently, must be removed [11]. Interactive image analysis can bypass the problem of debris and aggregates by selecting nuclei of interest, but Burger et al. [12] have underlined the bias then introduced. Moreover, this approach is very time consuming and unable to give statistically significant results in an acceptable delay for clinical oncology. To collect enough information, ICM-DNA must be automatised, especially for debris and aggregate elimination.

Automatic classification methods can be divided into two categories depending on whether they rely on individual methods or global methods.

Individual classification methods are based on measurement of parameters on objects to be sorted, derived [9] or not [21,26], from mathematical morphol-

ogy. Classification is done object per object, individually, from binary and/or grey level images, according to the value of parameters which characterise each element, using supervised methods (logistic regression [23], neural networks [13], multiparametric analysis [20]) or unsupervised methods (cluster analysis [16]).

Global methods do not use any individual measurement of parameters, but are based on the study of the whole image. These methods can be separated in two categories. The first one does not require any image transformation. The grey level image thresholding which allows the global sorting of bright objects versusdark ones is a good illustration. The second category is based on global morphological transformations of the whole image, characterised by the size and shape of the structuring element used. The thickness dependent sorting of objects, using erosion/reconstruction, is the simplest example which can be given. With this kind of transformation, objects remaining in the image belong to the category of the thickest elements, while objects which disappear belong to the category of the thinnest objects. The result of the classification is then included in the transformed image since the morphological transformation includes the sorting criterion. These global transformations can be performed on binary or grey level images and the computation time is independent of the number of objects per image.

The aim of the present study is to detail and compare the performance of two individual and one global automatic method applied to the elimination of debris and aggregates for DNA ploidy measurements. Interactive sorting (IS) reproducibility is also assessed.

2. Materials and methods

2.1. Biological material

The study was done on three archival tumours, one brain tumour, (a DNA aneuploid astrocytoma grade 2 according to Berner et al. [3]) and two carcinoma of the breast (no. 1 and no. 2, respectively DNA aneuploid and DNA diploid). The samples were prepared according to Van-Driel Kulker et al. [25] and DNA stained according to Feulgen and Rossenbeck [18], as described previously [17].

2.2. Image cytometry

The image cytometer consists of a BH2 Olympus microscope (OSI, France), a moving stage (Galai, Israel), a matrox PIP 1024 frame grabber (Matrox, Electronix systems, Ltd., Canada) and a Sony CCD camera (Sony, France). Segmentation of objects was achieved according to the Deriche algorithm [15]. Integrated optical density computations were done at a resolution of 512×512 pixels in 8 bits (1 pixel $= 0.11$ μm^2). The study was performed on the same set of stored images: 50 images for the brain tumour and 196 images for each breast carcinoma.

2.3. Automatic classification methods

2.3.1. Multiparametric analysis (MA) method
Automatic identification of nuclei, debris and aggregates was performed by the computation of 38 parameters for each segmented element [21] (11 parameters of size and shape, 18 statistic parameters calculated on original and edge-enhanced images and 9 texture parameters). Sorting was performed by reference to a knowledge base, specific of each tumour localisation, and obtained by interactive sorting (4291 normal nuclei and 1535 abnormal nuclei for breast carcinoma, 168 normal nuclei and 208 abnormal nuclei for brain tumours). Nuclei belonging to normal morphological categories were obtained from normal tissue of the corresponding localisation. Nuclei of abnormal morphology were obtained from tumours. Regardless of the localisation (brain or breast), the morphology of debris and aggregates was not memorized.

Automatic sorting refers to the representation of each group of objects by an ellipsoid in its own 38 parameter reference space. Each ellipsoid is rescaled to obtain the same unity value mean squared distance. To sort an unknown object, the Euclidean distance (Ed) from the object to the center of each rescaled ellipsoid is calculated. If Ed is superior or equal to two then the object is labelled as debris, or else the object is assigned to the closest group (shorthest Ed) [19]. Labelling is checked with respect to each parameter vs the calculated limits of the category (only the upper limits of the abnormal categories are not taken into account); if the check fails, the object is assigned to the next closest category or labelled as debris in case of second failure. Segmented events were classified in six categories of normal and abnormal nuclei, and one category of debris for the brain tumour. Segmented events were classified into 10 categories of normal and abnormal nuclei, and one category of debris for breast tumours [19].

2.3.2. Neural network (NN) method

Neural network classification was performed using the same 38 parameters and the same specific knowledge bases used for MA. The system used several NN of multilayer perceptrons type and was developed in the laboratory in C language. Each NN is devoted to the separation of one category from the other. Then for n categories, $n \times (n-1)/2$ parallel NN are required. Each NN has for input the 38 parameters and gives its output to a decision module, which produces the category of the element being used, refering to the outputs of the other NN. Using several parallel NN instead of a unique NN presents the advantage to automatically adapt the complexity of each NN to the difficulty of separating two categories. The number of categories used was strictly the same as for MA.

2.3.3. Mathematical morphology (MM) method

The method was developed to specifically eliminate small debris and aggregates [10]. This method is performed globally on images using mathematical morphology operators, without individual estimation of parameters. It is based on the use of size and intensity criteria for the elimination of small debris by successive top-hat transformations [22]. Concavity criterion based on the computation of watershed transformation [4] and a dodecagonal distance function leads to the elimination of aggregates. The method was adjusted by reference to learning of the morphology of a wide spectrum of small debris (307) and aggregates (38).

2.4. Comparison of performance of automatic methods by reference to an unique interactive sorting and for the three tumours

Because strict and precise elimination of debris and aggregates is required before DNA ploidy measurements [11], the aim of this comparison was only to test the ability of each method to characterise debris and aggregates vs undamaged nuclei on the same set of objects extracted from the same set of stored images. For each tumour, sorting of debris and aggregates was done interactively (two categories: undamaged nuclei versus debris and aggregates), in order to get a reference sorting. A total of 7120 elements was studied for the three tumours (3764 nuclei, 3356 debris and aggregates).

Performances of automatic methods were assessed using the computation of sensitivity (S) and false positive rates (FP). S is defined as the percentage of debris and aggregates correctly classified by automatic methods by reference to interactive sorting (IS). FP rate is defined as the percentage of undamaged nuclei misclassified as debris and aggregates. We defined a quality factor of sorting (Qt) expressed as follows:

$$Qt = S \times (100 - FP)/100.$$

Its varies from 0 to 100 (optimal sorting).

It should be noted that automatic methods were adjusted, in order to obtain the highest value of Qt.

Performances of automatic methods were also evaluated regarding the restitution of DNA ploidy abnormalities, after automatic elimination of debris and aggregates. For this purpose, DNA ploidy histograms were calculated after IS and each automatic sorting. Five indices describing DNA ploidy abnormalities were evaluated for each histogram. Four were chosen according to the recommendations of the consensus on image DNA cytometry [7]: 5c exceeding rate (5cER), 2c deviation index (2cDI) (per Böcking et al. [5]), DNA malignancy grade (DNA MG) (per Böcking and Auffermann [6]) and distribution entropy (DE) (per Stenkvist and Strande [24]). ICM-DNA data obtained were also post-processed by MCycleAV® (Phoenix Flow Systems, San Diego, USA) for S phase fraction estimation. As advised by Berger et al. [1,2], the S phase model used was the zero order one [14].

In order to obtain a comprehensive comparison, values were normalised for each DNA index with respect to reference values (set to 100) obtained after IS (when reference values were equal to zero, values obtained for automatic methods were not calculated).

For each automatic method and each tumour, the percentage of error for the calculation of each DNA index, was computed using the following formula:

$$error = \left| 100 \times \left(\frac{n}{p} \right) - 100 \right|,$$

where n is the value of the DNA index obtained after automatic sorting (before normalization) and p the reference value of the DNA index obtained after IS (before normalization). The error calculated for each index was then calculated for each automatic method and for the three tumours. The mean error of the calculation of the whole index was also evaluated for the three tumours and for each automatic method.

2.5. Comparison of performance of six interactive sortings and of the three automatic methods for astrocytoma grade 2

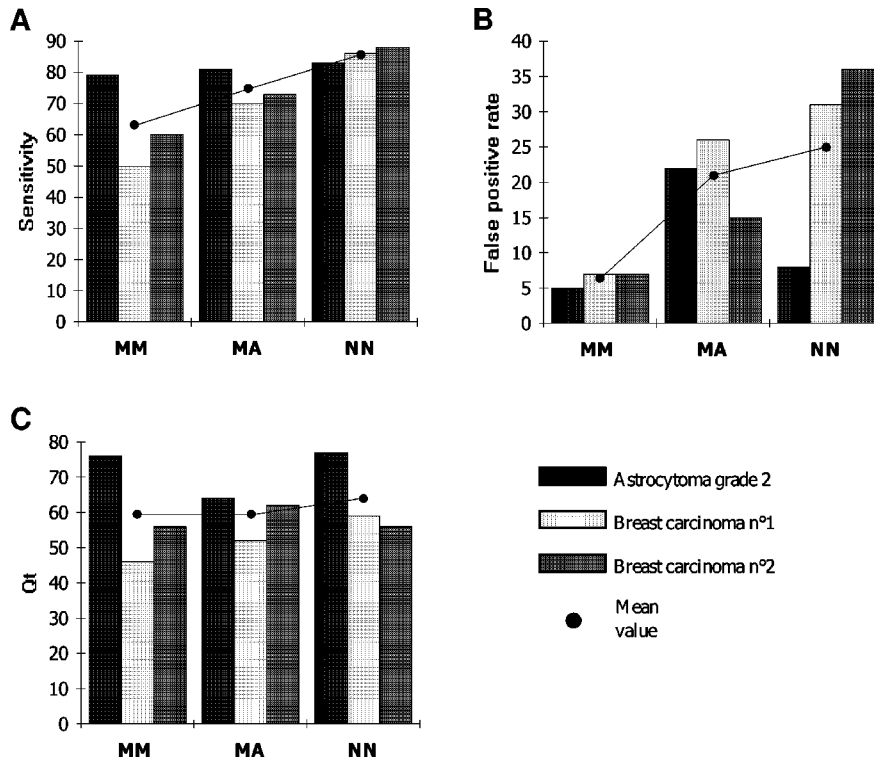In order to evaluate, as a comparison, the reproducibility of interactive elimination of debris and ag-

Fig. 1. Sensitivities (A), false positive rates (B) and quality factor of sorting (C) for the three cases and for the three automatic methods. Mean values for the three cases are represented as black dots for each method.

gregates for DNA ploidy abnormality estimation, six IS of the same set of images of astrocytoma grade 2 were obtained by the same operator. Three IS (nos 1–3) were done considering only two categories: undamaged nuclei versus debris and aggregates. Three other IS (nos 4–6) were done considering seven categories: six categories of normal and abnormal nuclei versus one category of debris and aggregates [10]. DNA ploidy histograms were calculated and the 5 DNA indices were calculated only for nuclei (one category for IS nos 1–3, six pooled categories for IS nos 4–6).

DNA indices were calculated after elimination of debris and aggregates by the three automatic methods and compared to those obtained by the six interactive sortings.

3. Results

3.1. Comparison of the performances of the three automatic methods by reference to a unique interactive sorting and for the three tumours

Sensitivities (S), false positive (FP) rates and optimal quality factor Qt are given in Fig. 1 for the three

methods. Whatever the tumour studied, the sensitivities obtained with NN method are stable and are the highest (mean = 85%) as compared to MM method (mean = 63%) and MA method (mean = 75%) (Fig. 1A). Concerning FP rates, MM gives stable and low values (mean = 6%) whereas values obtained with NN and MA methods vary to a large extent (Fig. 1B). The mean values of Qt obtained for the three cancer cases are quite similar for the three methods. For an equivalent value of Qt, one must notice that sensitivities and FP rates obtained with MM method are the lowest, whereas sensitivities and FP rates obtained with NN method are the highest.

Performances of automatic methods, evaluated with respect to the restitution of DNA ploidy abnormalities, are presented in Fig. 2. Only one DNA indice (DE) was always correctly evaluated by the three methods and for the three tumours (Fig. 2D). The fact that DE remains unchanged suggests that the misclassification of debris and aggregates by the three automatic methods has no effect on the distribution of the information content. That is to say, that the overestima-
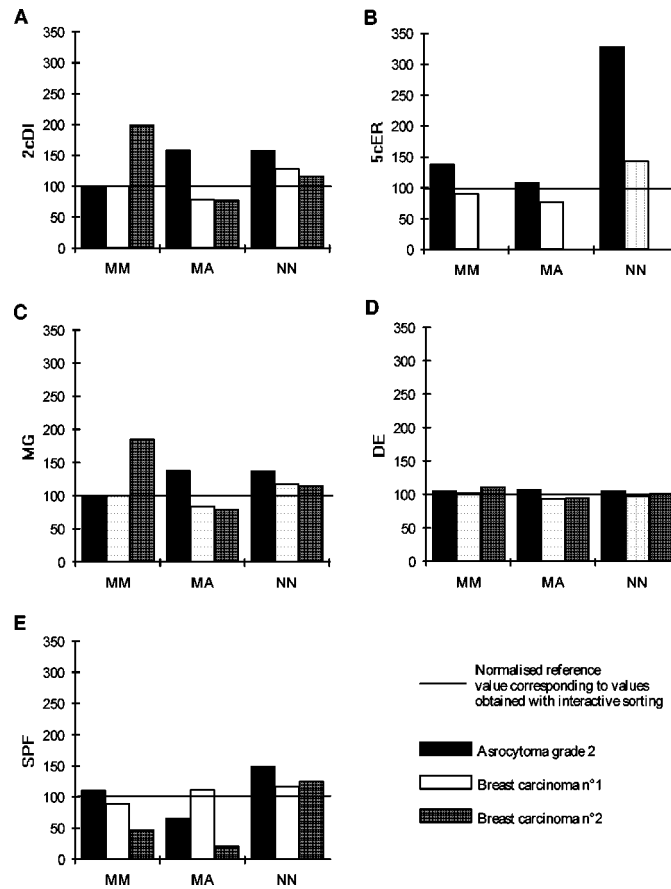
Fig. 2. 2cDI (A), 5cER (B), MG (C), DE (D), SPF (E) normalised values obtained after the elimination of debris and aggregates by each of three automatic methods.

tion of object number (misclassified debris) or the underestimation (misclassified nuclei) occurs in an homogeneous pattern. This observation shows that using one of the proposed automatic methods generates a systematic error on the information content carried by sorted data, which, consequently occurs on the parameters computed from this information. The values obtained for the other DNA indices are more variable, especially SPF values (Fig. 2E). Apart from DE, the MA and NN methods lead to more variable evaluation of DNA ploidy indices than the MM method. The mean error obtained with MM for each DNA indice and for the three tumours is more often low (less than or about 20%) as opposed to the MA and NN methods (Fig. 3A–3E). The mean error for the five DNA indices and the three tumours is about 27% for the MA method and 65% for the NN method as compared to 21% for the MM method (Fig. 3F).
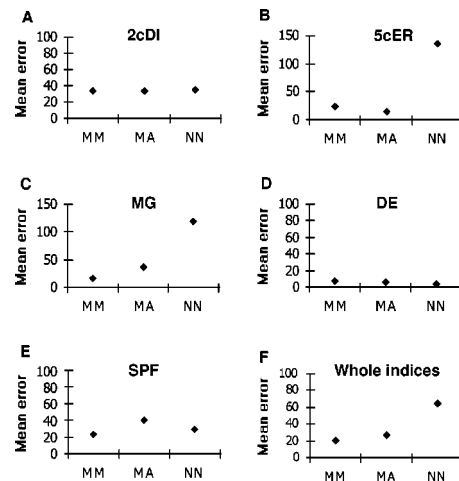

Fig. 3. Mean errors for three automatic methods on each DNA indice evaluation for the three tumours: 2cDI (A), 5cER (B), MG (C), DE (D), SPF (E). Mean error occurring on the five indices taken as a whole for the three tumours is also represented (F).
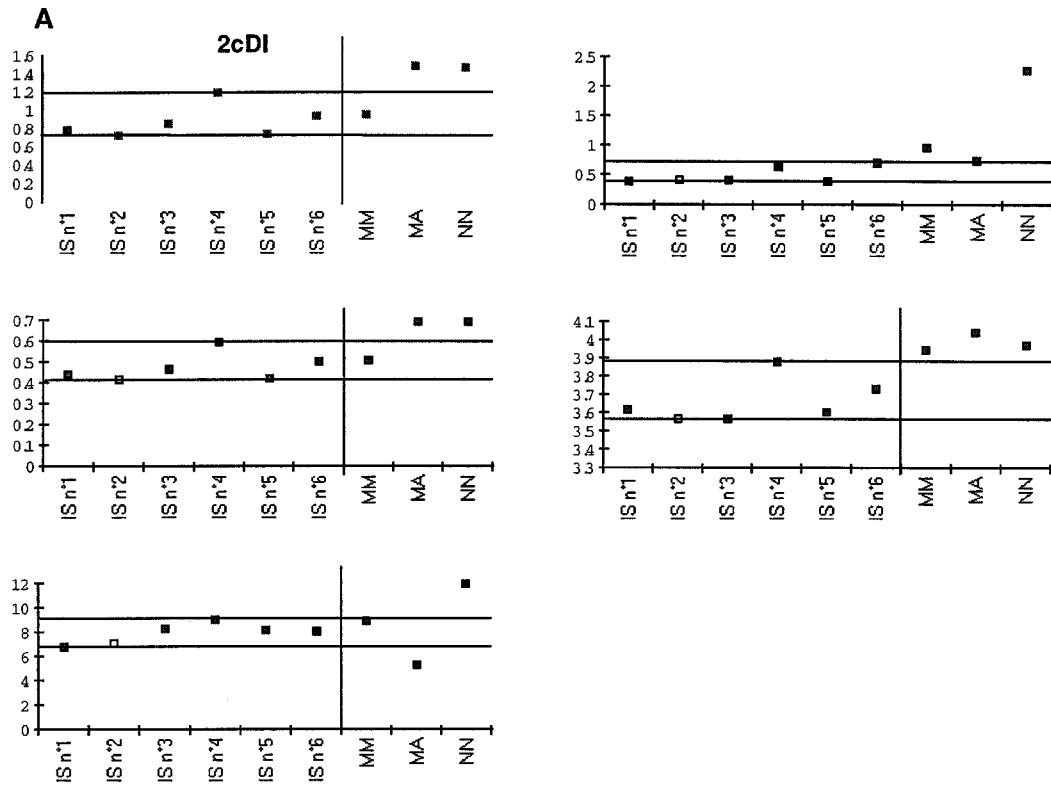
Fig. 4. DNA indices values (A: 2cDI, B: 5cER, C: MG, D: DE, E: SPF) obtained after the elimination of debris and aggregates by the six interactive sortings (on the left) and by the 3 automatic methods (on the right) for astrocytoma grade 2. Horizontal lines delimit lowest and highest values for each DNA indice obtained after the interactive elimination of debris and and aggregates. Vertical lines separate interactive sortings from automatic sortings.

3.2. Comparison of the performances of six interactive sortings and of the performances three automatic methods in terms of restitution of DNA ploidy abnormalities for astrocytoma two

Comparison of the performances of six IS is illustrated in Fig. 4 (on the left) for astrocytoma grade two. DNA ploidy indices show considerable variation, especially for IS nos 4 and 6. It should be noticed that the restitution of DNA ploidy abnormalities is more variable than seven categories for IS nos 4–6 (six categories of nuclei and one of debris and aggregates) than when using two categories for IS nos 1–3 (one category of nuclei and one of debris and aggregates). Comparison of performances of the three automatic methods is illustrated in Fig. 4 (on the right) for astrocytoma grade two. DNA ploidy indices also exhibit considerable variation. Nevertheless, differences have been observed between performances of automatic methods. DNA indices obtained after automatic elimination of

debris and aggregates by MM (Fig. 4) lead globally to values included between the lowest and highest values obtained with interactive sorting (Fig. 4). MA and NN automatic methods lead globally to more variable DNA index values.

4. Discussion

Careful elimination of numerous debris and aggregates (which can represent an average of half the events using this type of sample preparation [11]) is one of the prerequisites for a reliable estimation of DNA ploidy abnormalities [11]. In order to provide pathologists with alternative tools to tedious interactive elimination of debris and aggregates, three different automatic classification methods were tested here. Results obtained show that although the three methods tested can globally exhibit the same quality factor for sorting (Qt), sensitivity and false positive rates are quite different, and these differences have consequences for the quality of the restitution of DNA ploidy abnormalities.

Good quality of DNA ploidy abnormality evaluation is the final aim of sorting and must guide the choice and refinement of automatic sorting methods. At best, and for the three tumours studied, the MM method gives a global estimation of the five DNA indices with a mean error of 23%. This underlines the inadequacy of Qt to evaluate the reliability of a sorting method. The false positive rate seems to be the better indicator for the estimation of sorting quality. For this application, a good classification method must favour power against false positive detection, rather than sensitivity.

The concomittant study of the performances of six interactive sortings (IS) has shown a surprising degree of unreproducibility which seems to be function of the complexity of sorting (number of classes considered). This raises the question of the use of an interactive sorting method as a reference for comparing automatic methods [8].

Compared to DNA index values obtained after interactive elimination of debris and aggregates, automatic methods globally give neighbouring values. The MM based method gives the best results and, to a lesser extent, the multiparametric analysis based method (MA). It seems then that global and non parametric classification methods such as the MM method can be successfully compete with the more popular parametric methods (MA and NN). Nevertheless, performances of MA and NN could probably be improved by searching and eliminating redundant or useless parameters from among the 38 parameters used here (work in progress in our laboratory). In conclusion, automatic classification methods can offer on alternative to tedious interactive elimination of debris and aggregates for DNA ploidy measurements of archival tumours. In particular, global mathematical morphology based methods seems to be promising.

Acknowledgements

References

[1] E. Bergers, R. Montironi, P.J. van Diest, E. Prete and J.P.A. Baak, Inter-laboratory reproducibility of semi-automated cell cycle analysis of flow cytometric DNA-histrograms obtained from fresh material of 1295 breast cancer cases, *Hum. Pathol.* **27** (1996), 553–560.

[2] E. Bergers, P.J. van Diest and J.P.A. Baak, Reproducibility of semi-automated cell cycle analysis of flow cytometric DNA-histograms of fresh breast cancer material, *Analyt. Cell. Pathol.* **8** (1995), 1–13.

[3] A. Berner, H. Danielsen, N. Juul, M. Juul, E. Pettersen and A. Reith, Caveats in estimation of DNA-ploidy in paraffin embedded specimens of primary prostate cancer and lymph node metastases by flow cytometry and image cytometry, *Analyt. Cell. Pathol.* **5** (1993), 339–352.

[4] S. Beucher and C. Lantuejoul, Use of watersheds in contour detection real-time edge and motion detection/estimation, in: *Proc. Int. Whorshop on Image Process*, Rennes, 1979.

[5] A. Böcking, C.P. Adler, H.H. Common, M. Hilgarth, B. Granzen and W. Auffermann, Algorithm for a DNA-cytophotometric diagnosis and grading of malignancy, *Analyt. Quant. Cytol.* **6** (1984), 1–8.

[6] A. Böcking and W. Auffermann, Algorithm for DNA cytophotometric diagnosis and grading of malignancy, *Analyt. Quant. Cytol.* **8** (1986), 363.

[7] A. Böcking, F. Giroud and A. Reith, Consensus Report of the European Society for Analytical Cellular Pathology: Task force on standardization of diagnostic DNA image cytometry, *Anal. Quant. Cytol. Histol.* **17** (1995), 1–7.

[8] C. Boudry, Classification cellulaire par morphologie cellulaire. Thèse de Doctorat of the University of Caen, 1997.

[9] C. Boudry, P. Herlin, M. Coster, J.L. Chermant, B. Sola and M. Henry-Amar, Using mathematical morphology to eliminate debris and aggregates before DNA ploidy measurement of solid tumors, *Acta Stereol.* **14** (1995), 121–128.

[10] C. Boudry, M. Coster, P. Herlin, B. Sola and J.L. Chermant, Global and non parametric classification methods using mathematical morphology: application to DNA ploidy measurement of archival tumors, *Micros. Microanal. Microstruct.* **7** (1996), 477–485.

[11] C. Boudry, P. Herlin, M. Coster, B. Sola and J.L. Chermant, Influence of debris and aggregates on image cytometry DNA measurement of archival tumors, *Anal. Quant. Cytol. Histol.* **19** (1997), 153–157.

[12] G. Burger, M. Aubele, U. Jutting and G. Auer, Interactive cytometry: Chance or evil of bias, *Path. Res. Pract.* **188** (1992), 391–395.

[13] E. Davalo and P. Naim, *Des Réseaux de Neurones*, Eyrolles, 1990.

[14] P.N. Dean and J.H. Jett, Mathematical analysis of DNA distribution derived from flow microfluorometry, *J. Cell. Biol.* **60** (1974), 523–527.

[15] R. Deriche, Using Canny's criteria to derive a recursively implemented optimal edge detector, 1987, pp. 167–187.

[16] E. Diday, J. Lemaire, J. Pouget and F. Testu, *Eléments d'Analyse et de Données*, Dunod, Paris, 1982.

[17] F. Duigou, I. Galle, P. Herlin and A.M. Mandard, Improvement in slide preparation from archival material for automated DNA measurement by image analysis, *Anal. Quant. Cytol. Histol.* **19** (1997), 167–173.

[18] R. Feulgen and H. Rossenbeck, Mikroskopich chemischer nacheis einer nucleinsaure von typus der thymonuckleinsaure und die darauf beruhende elektive farbunf von zellkernen in mikroskopichen praparaten, *Z. Physiol. Chem.* **135** (1924), 203–224.

[19] P. Herlin, E. Masson, F. Duigou, B. Plancoulaine, J.P. Signolle, A.M. Mandard, F. Angot, D. Deman, P. Belhomme, J.B. Joret, T. Datry, O. Rougereau and D. Bloyet, Fully automated quantitative analysis of DNA in solid tumors, *Bull. Cancer* **84** (1997), 685–692.

[20] L. Lebart, A. Morineau, J.P. Fenelon, *Traitement des Données Statistiques. Méthodes et Programmes,* Dunod, 1982.

[21] E. Masson, P. Herlin, I. Galle, F. Duigou, P. Belhomme, D. Bloyet and A.M. Mandard, Automatic classification of cellular elements of solids tumors: application to DNA quantitation, *Acta Stereol.* **13** (1994), 75–81.

[22] F. Meyer, Contrast features extraction. Quantitative analysis of microstructures in Material science, *Biology and Medicine*, Dr Riederer Verlag (1977), 374–380.

[23] M. Pagano and K. Gauvreau, *Principles of Biostatistics*, Duxbury Press, 1993.

[24] B. Stenkvist and G. Strande, Entropy as an algorithm for the statistical description of DNA cytometric data obtained by image analysis microscopy, *Analyt. Cell. Pathol.* **2** (1990), 159–165.

[25] A.M.J. Van Driel-Kulker, W.E. Mesker, M.J.M. Van der Burg and J.S. Ploem, Preparation of cells from paraffin-embedded tissue for cytometry and cytomorphologic evaluation, *Analyt. Cell. Pathol.* **9** (1987), 225–231.

[26] L. Wheeless and R. Robinson, Classification of red blood cells as normal, sickle, or other abnormal, using a single image analysis feature, *Cytometry* **17** (1994), 159–166.