Detecting Overlapped Nuclei Regions in the Feulgen-Stained Cytological Smears

Bogusław D. Piętka and Annamonika Dulewicz

Abstract. Projects related to clinical implementation of computerized image processing in cytology face a common problem of distinguishing between artifacts and the objects of interest which should be measured and analyzed. Secondary Screening Instruments are computerized devices used to detect possible presence of cancerous or precancerous cells and, independently or in collaboration with pathologist, select slides for additional manual review. In order to do the work efficiently they have to detect actual abnormalities distinguishing them from artifacts in the form of overlapping cells or nuclei. The paper reports a trial approach to perform this important discrimination.

1 Introduction

The potential benefit of Secondary Screening Instruments (SSI) is that they increase the overall sensitivity of a cytological screening by detecting the presence of possible abnormalities missed during a primary screening. Be-cause primary screening (i.e. first time examination of a slide) is not 100

Most projects related to clinical implementation of computerized image processing in cytology face a common problem of distinguishing between artifacts and the objects of interest which should be measured and analyzed. Since most of algorithms for cancer cell identification rely on some kind of abnormality detection, artifacts left in a sample would generate too many undesired, false-positive alarms, making such a system impractical.

Bogusław D. Piętka · Annamonika Dulewicz
Laboratory of Fundamentals of Computer-Aided Image Diagnostics,
Institute of Biocybernetics and Biomedical Engineering, Polish Academy of Sciences,
ul. Ks. Trojdena 4, 02-109 Warsaw, Poland
e-mail: bpietka@ibib.waw.pl

R. Burduk et al. (Eds.): *CORES 2013*, AISC 226, pp. 621–628.
DOI: 10.1007/978-3-319-00969-8_61

Generally, artifacts are undesired objects or phenomena influencing the appearance of a smear and obstructing, or even preventing, proper analysis of important factors of cytological sample. Several examples of the artifacts are presented in Fig. 1.1.

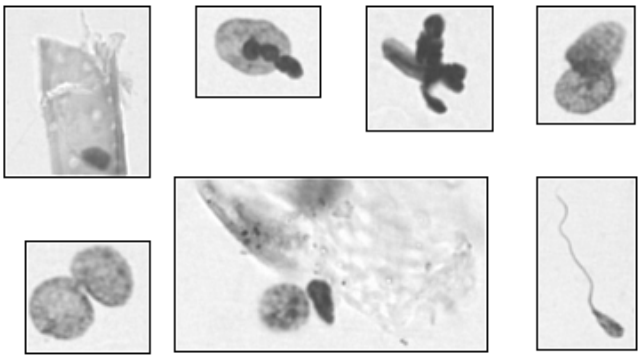

Fig. 1.1 Examples of the artifacts found in a cytological smear (Feulgen-stained urine sample

Although there are different types of artifacts, in this paper we are focusing on just one, particular sort of them; two overlapping nuclei (Fig. 1.2).

Fig. 1.2 Examples of particular type of artifacts in the form of two normal (non-cancerous) overlapping nuclei

2 Searching for Textural Markers of the Overlap

In the case of evident nuclei overlap it is relatively easy for a human ob-server to distinguish two regions of different textural nature (Fig. 2.1). Let us try to find some objective, quantitative indicators of the occlusion. There are three basic methods used to describe the texture of a region: statistical, structural and spectral. Since the structural and spectral techniques are pri-marily used to find a kind of periodicity, directionality or repetitive pixel arrangement, they are not well suited for the problem under consideration. Therefore, the only promising approach seems to be statistical one. Starting from the simplest, first order measures of a region texture,

the his-tograms h(zi) (i = 1, 2, … N, zi = 0, 1 … N-1, N = 256) of the two distinct areas (AuB and C) from figure 2.1 are computed (Fig.2.2). Actually, prior to histogram calculation, all images are normalized by means of the contrast stretching to saturate the bright background (255).

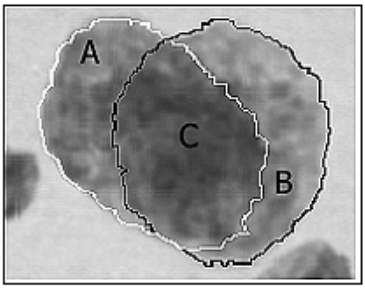

Fig. 2.1 An example of manually selected regions used for studying the overlap phenomenon

In general, intensity distributions of regions A and B are computed sepa-rately. Since our experiments proved they are highly similar, it was decided to combine them into one region. Thus, we get two histograms describing one, overlapping object:

$$H_{A_UB} = H_A + H_B \text{ and } H_C$$

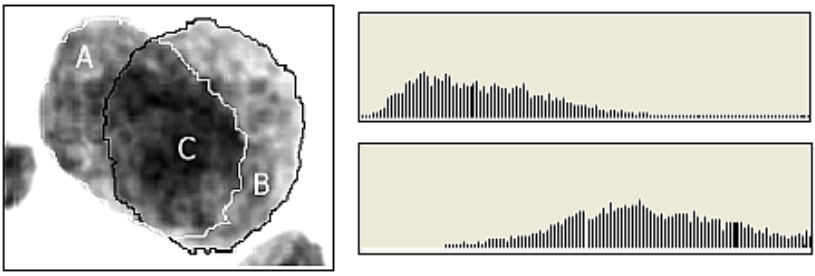

Fig. 2.2 The image from Fig.2.1 after normalization and grey level distributions of two distinct areas: overlapped (C-top) and non-overlapped (AuB-bottom)

Although differences between locations of the two histograms on the grey level axes are clearly visible, they do not constitute sufficient overlay indica-tion. They are not even the most essential for the overlap identification. For example, a single malignant nucleus also tends to have its histogram shifted to the darker grey levels as a result of increased presence of heterochromatin (Fig. 2.3). However, we have noticed statistical significance of the higher order central moments of the histograms in distinguishing such cases from the overlaps.

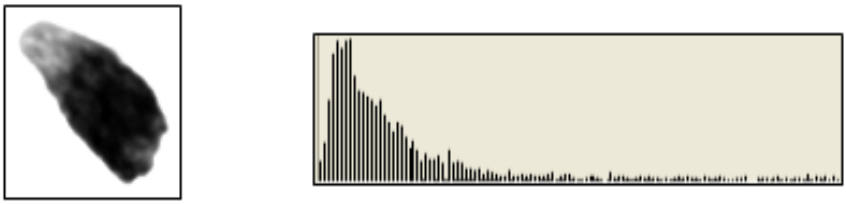

Fig. 2.3 Suspicious nucleus and the grey-level distribution of its dark region

To construct objective statistical descriptors of a texture, four parameters of the grey-level distribution were computed. In addition to the mean value μ they were: variance, skewness and kurtosis according to the following formulas:

$$\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}[h(z_i)-\mu]^2$$

$$\gamma = \frac{1}{(N-1)\cdot\sigma^3}\sum_{i=1}^{N}[h(z_i)-\mu]^3$$

$$\delta = \frac{1}{(N-1)\cdot\sigma^4}\sum_{i=1}^{N}[h(z_i)-\mu]^4$$

where the mean value μ is defined as

$$\mu = \frac{1}{N}\sum_{i=1}^{N}[h(z_i)\cdot z_i]$$

and N = 256 is the total number of grey levels in the histogram.

Fifty image samples for each of the three area types (normal, dark-malignant and overlapped) were collected, processed and the resulting data supplied to unsupervised clustering algorithm. Our goal was not to perform precise discriminant analysis but to discover any general patterns that may be present in the collected data set.

3 Experimental Results

In the very beginning of the data analysis process it was found that the mean value of the histograms has the lowest discrimination power in comparison with others. Therefore, it was decided to perform unsupervised clustering of the data basing only on the three descriptors: variance, skewness and kurto-sis. Moreover, since the goal of the investigations was to get some qualita-tive knowledge about the phenomenon

we are going to show the results graphically. For this purpose the KlustaWin software package was used that is able to perform different kinds of data clustering and present the results as a two or three-dimensional scatter-plots [5]. The software is an open-source project for scientific purposes.

It may be informative to see the scatter-plots from three different perspec-tives (Fig. 3.1-3.3) where selected histogram parameters were assigned to the X, Y, Z axes as follows

$$\sigma^2 \implies X \text{ (variance)}$$

$$\gamma \implies Y \text{ (skewness)}$$

$$\delta \implies Z \text{ (kurtosis)}$$

Numerical values of the data on every axis were normalized to the same range 0-1.

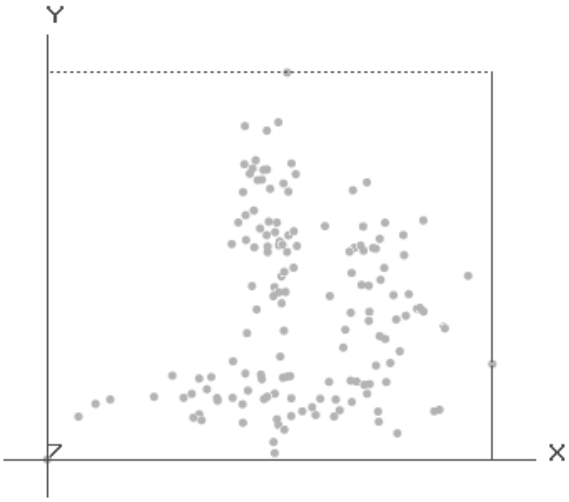

Fig. 3.1 2-D scatter-plot as can be seen from variance-skewness surface perspective

From the perspective of variance-skewness surface it is impossible to clearly distinguish any clusters. However, perceptive observer would draw a divi-sion line orthogonal to X axis at approximately x=0.55. The zones created in this way represent overlapping textures (left) and non-overlapped regions (right).

It may be easily noted that the scatter-plot from Fig. 3.2 (variance-kurtosis) demonstrates much better separation than that from Fig. 3.1 (variance-skewness), and mainly thanks to the kurtosis (Z axis). This surface is best suited for distinguishing between dark-malignant regions and overlapped regions which usually have high density as well. Finally, the last scatter-plot in Fig. 3.3 is definitely the best one. It seems that the pair of descriptors skewness-kurtosis would be enough to

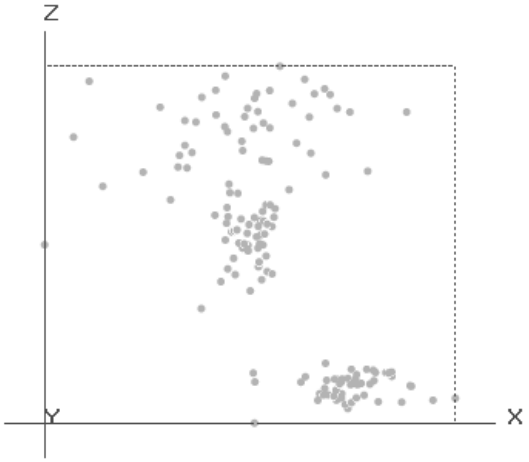

Fig. 3.2 2-D scatter-plot as can be seen from variance-kurtosis surface perspective

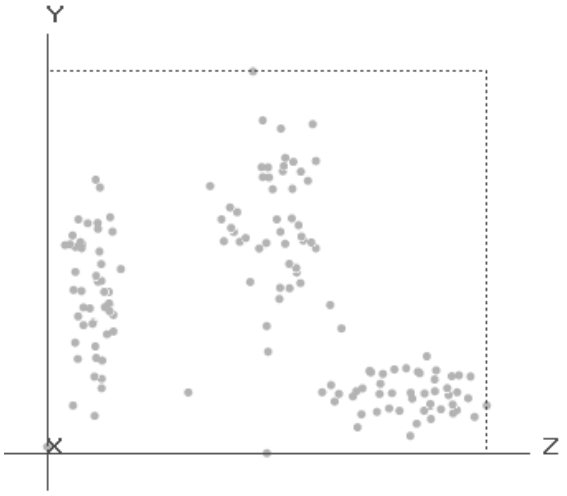

Fig. 3.3 2-D scatter-plot as can be seen from kurtosis-skewness surface perspective

obtain very good separation between the three kinds of textures under study. Suspicious textures are characterized by the high kurtosis (cluster on the right). On the other hand, non-overlapped textures histograms have wide tails implicating low values of the kurtosis. To get an impression of the data distribution in full three-dimensional space there is a 3-D scatter-plot shown in Fig. 3.4 below.

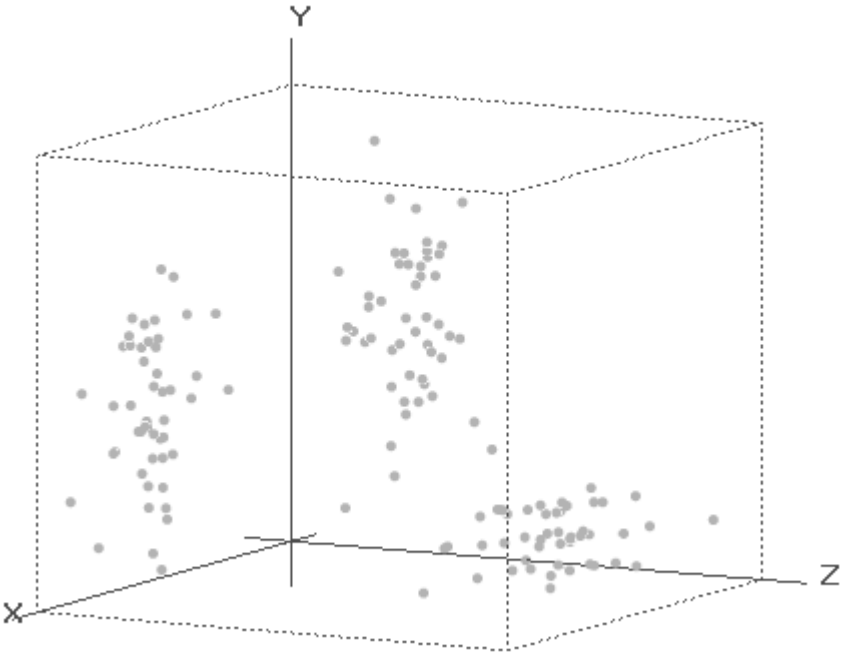

Fig. 3.4 3-D scatter-plot combining all three parameters; variance, skewness and kurtosis

4 Summary

It was found useful when describing textural features of nuclear overlaps to divide the problem into two distinct levels of complexity. First, in every ini-tially extracted object, a trial is performed to divide the total area into two regions with different textural properties. If the division may be performed successfully, a detailed, histogram-based analysis of the textures is conducted to identify overlapping regions.

Since it was just an initial approach to solve the problem, the regions were extracted manually and then subjected to computer analysis by means of image processing tools. It was shown that combination of properly selected descriptors of the region texture (skewness and kurtosis) can perform quite good separa-tion in the data space. Simple reclassification based on the centroids of the clusters (Fig. 3.4) is presented in the table below (Fig. 4.1).

Unfortunately, at this particular moment we have not finalized the task of designing a fast and reliable procedure for automatic segmentation of the whole object which is necessary to complete the task. The work is under progress and will be the subject of a paper in near future.

Expert→	Non-Overlapped	Overlapped	Cancerous
Non-overlapped	50	3	0
Overlapped	0	45	0
Cancerous	0	2	50

Fig. 4.1 Results of the sample textures reclassification (150 items)

References

1. Ravishankar, A., Lohse, G.: Identifying high level features of texture perception. CVGIP: Graph. Models Image Process. 55(3), 218–233 (1993)
2. Haralick, R.M.: Statistical and Structural Approaches to Texture. In: van Leeuwen, J. (ed.) Computer Science Today. LNCS, vol. 1000, pp. 786–804. Springer, Heidelberg (1995)
3. Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural Features for Image Classification. IEEE Transactions on Systems, Man and Cybernetics SMC-3(6), 610–620 (1973)
4. Harris, K.: (Rutgers University, USA), Heitler B. (University of St Andrews, UK), KlustaWin - A program for unsupervised classification of multidimensional continuous data, `http://www.st-andrews.ac.uk/~wjh/klustawin/`