ADVANCED STATISTICAL MATRICES FOR TEXTURE CHARACTERIZATION: APPLICATION TO DNA CHROMATIN AND MICROTUBULE NETWORK CLASSIFICATION

Guillaume Thibault, Jesús Angulo, Fernand Meyer

CMM-Centre de Morphologie Mathématique, Mathématiques et Systèmes, MINES ParisTech
35, rue Saint Honoré, 77305 Fontainebleau Cedex, France

ABSTRACT

This paper presents significant improvements of Gray Level Size Zone Matrix (GLSZM) which is a bivariate statistical representation of texture, based on the co-occurrences of size/intensity of each flat zone (connected pixels of the same gray level). The first improvement is a multi-scale extension of the matrix which merges various quantizations of gray levels. A second alternative is proposed to take into account radial distribution of zone intensities. The third variant is a generalization of the matrix structure which allows to analyze fibrous textures, by changing the pair intensity/size for the pair length/orientation of each region. The interest of these improved descriptors is illustrated by texture classification problems arising from quantitative cell biology.

Index Terms— Texture Characterization, Structural Statistical Matrices, Gray Level Size Zone Matrix (GLSZM).

1. INTRODUCTION

Texture characterization and classification [1] is one of the fundamental task in low-level computer vision. Two main families of approaches appear highlighted in the state-of-the-art: based on statistical analysis [2, 3, 4] and based on wavelets descriptors [5].

Our starting point is the original notion of Gray Level Size Zone Matrix (GLSZM), based on the co-occurrences of size/intensity of each flat zone (connected pixels of the same gray level), which was introduced recently in [6] as an alternative to the joint gray level-run length distribution [3, 4]. More precisely, the aim of this paper is to offer some alternative bivariate statistical representations of texture. The first improvement, discussed in Section 2, is a multi-scale extension of the matrix which merges various quantizations of gray levels and which avoids selecting the optimal quantization. A second alternative is proposed in Section 3 to take radial distribution of zone intensities into account. The third variant, studied in Section 4 is a generalization of the matrix structure which allows to analyze fibrous textures, by changing the pair intensity/size for the pair length/orientation of each region. The interest of these improved descriptors is illustrated by texture classification problems arising from quantitative cell biology in section 5.

2. MULTIPLE GRAY LEVEL SIZE ZONE MATRIX

Let $f(\mathbf{x}) : E \rightarrow \mathcal{T}$ be a gray-level image, where $E \subset \mathbf{Z}^2$ is the space pixels $\mathbf{x} \in E$ and the image intensities are discrete values which range in a closed set $\mathcal{T} = \{t_1, t_2, ..., t_N\}$, $\Delta t = t_{i+1} - t_i$, e.g., for a 8 bits image we have $t_1 = 1$, $N = 256$ and $\Delta t = 1$. Let us assume also that image f is segmented into its J flat zones (i.e., connected regions of constant value): $E = \cup_{j=1}^{J} R_j[f]$, $\cap_{j=1}^{J} R_j[f] = \emptyset$. The size (surface area) of each region is $s(j) = |R_j[f]|$. Hence, we consider that each zone $R_j[f]$ has associated a constant gray-level intensity $g(j)$.

2.1. Reminder on Gray Level Size Zone Matrix (GLSZM)

The GLSZM of a texture image f, denoted $\mathcal{GS}_f(s_n, g_m)$, provides a statistical representation by the estimation of a bivariate conditional probability density function of the image distribution values. It is calculated according to the pioneering Run Length Matrix principle [3]: the value of the matrix $\mathcal{GS}_f(s_n, g_m)$ is equal to the number of zones of size s_n and of gray level g_m. The resulting matrix has a fixed number of lines equal to N, the number of gray levels, and a dynamic number of columns, determined by the size of the largest zone as well as the size quantization. The more homogeneous the texture, the wider and flatter the matrix. More precisely, we can calculate all the second-order moments of $\mathcal{GS}_f(s_n, g_m)$ as compact texture features [4]. Figure 1 shows an example of the calculation of such a matrix.

1	2	3	4
1	3	4	4
3	2	2	2
4	1	4	1

(a)

Level g_m	*Size zone*, s_n 1	2	3
1	2	1	0
2	1	0	1
3	0	0	1
4	2	0	1

(b)

Fig. 1. Example of the GLSZM filling for an image texture of size 4×4 with 4 gray levels.

This matrix does not require calculations in several directions, contrary to Run Length Matrix (RLM) [3] and Co-occurrences Matrix (COM) [7]. RLM and COM are appropriate for periodic textures whereas the GLSZM is typically adapted to describe heterogeneous non periodic textures. In addition, due to the intrinsic segmentation, texture description in GLSZM is more regional than the point-wise-based representation of COM. However, it has been empirically proved that the degree of gray level quantization still has an important impact on the texture classification performance. For a general application it is usually required to test several gray level quantization in order to find the optimal one with respect to a training dataset.

2.2. Multiple gray level quantization of GLSZM

Instead of optimizing the number of gray levels N, we propose to construct a multiple scheme with various matrices and then to combine them into a single matrix. The principle of the Multi Gray Level Size Zone Matrix (MGLSZM), for an image of 8 bits, is to calculate 8 GLSZM for 8 different numbers of gray levels: $N_k = 2^k$, $k = 1, 2, ..., 8$, and to merge the resulting matrices by a weighted average:

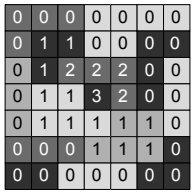
$$\widetilde{\mathcal{GS}}_f(s_n, g_m) = \frac{1}{\sum_k w_k} \sum_{k=1}^{8} w_k \mathcal{GS}_f^{N_k}(s_n, g_m)$$

where $\mathcal{GS}_f^{N_k}(s_n, g_m)$ is the GLSZM of image f calculated from $\mathcal{T}$ quantized in N_k gray levels. Weights distribution in MGLSZM is given by a gaussian function centered between $N_4 = 16$ and $N_5 = 32$ gray levels: this distribution penalizes extreme values of the number of gray levels because low levels contain limited structural information and high levels are sensitive to noise. By the way, the weights could be automatically learnt or adapted a priori for a specific application. For different values of N_k, the matrices $\mathcal{GS}_f^{N_k}(s_n, g_m)$ have different dimensions. Even if we consider that the size of regions is quantized in the same intervals (same number of columns), the number of rows is equal to N_k. To solve this drawback, we propose to replicate each of the N_k rows in order to finally obtain 256 rows.

The different gray-level quantization can also be interpreted as a segmentation into λ-flat zones [8], where the value of λ is associated to the corresponding $\Delta t = 256/N_k$. From a computational viewpoint, the multiple GLSZM $\widetilde{\mathcal{GS}}_f(s_n, g_m)$ requires to fill 8 individual GLSZM, but it is generally more efficient for texture classification (see application Section 5).

3. GRAY LEVEL DISTANCE-TO-BORDER ZONE MATRIX

Texture in natural objects is often non stationary in the space; for instance, the texture can vary radially with respect to the center of the object. Such is the case in one of the applications considered in Section 5. We have to characterize the DNA organization (i.e., chromatin texture) in cell nuclei, see Fig. 5. More precisely, the "quantity of DNA" is represented by the intensity of the pixel (i.e., the higher the pixel gray level, the higher quantity of DNA). As we observe from the examples, the distribution of DNA is not stationary and, for some classes, it is usually closer to the center of the nuclei. Then, in order to characterize such radial textures, we propose a descriptor named Gray Level Distance-to-border Zone Matrix (GLDZM), denoted $\mathcal{GD}_f^{N_k}(d_n, g_m)$. The new statistical matrix yields the number of zones of intensity g_m at a distance of d_n from the border of support space E^c (shortest Euclidean distance, see example of Fig. 2).



(a)

Level g_m	*Distance zone*, d_n 0	1	2	3
1	3	1	0	0
2	1	1	0	0
3	2	0	1	0
4	3	1	0	1

(b)

Fig. 2. Example of a texture with four gray levels, where each flat zone is valued with 4-connectivity distance to the border E^c (in white) (a) and the resulting GLDZM (b).

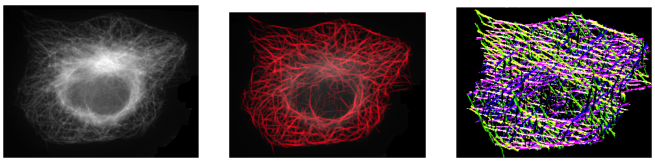
In practice, to accelerate computation time, the distance function is computed for the whole support space of the texture: $D(\mathbf{x}, E) = \inf\{d(\mathbf{x}, \mathbf{y}),\ \mathbf{y} \in E^c\}$, where $d(\mathbf{x}, \mathbf{y})$ is typically obtained using a discrete metric approximation to the Euclidean distance. Then, for each region $R_j[f]$, the corresponding value of distance is obtained as its smallest value in the distance map: $d_j = \inf\{D(\mathbf{z}, E),\ \mathbf{z} \in R_j[f]\}$. It is obvious how to generalize this matrix to construct a Multi Gray Level Distance-to-border Zone Matrix $\widetilde{\mathcal{GD}}_f(s_n, d_m)$.

4. ORIENTATION LENGTH ZONE MATRIX

The GLSZM and GLDZM are statistical descriptors which assume that the texture is composed of a random non periodic tiling of homogeneous zones, each one being described by its intensity value and its size/distance-to-border. This principle, which is appropriate to describe intensity-dependent homogeneous vs. heterogeneous textures [6], is not compatible with other kinds of more structured textures; in particular, with fibrous textures. Let us consider the example of microtubule network given in Fig. 3(a), which is another case studied in Section 5. Similar texture images can be found in other natural objects such as wood, carbon, wool, etc. On the one hand, we observe that the intensity of the fiber is not an important feature. Fibers can mainly be described by their length, width variation, orientation and tortuosity [9]. We assume here that our fibrous texture is a random network of thin fibers

of limited width variation and low tortuosity. On the other hand, the flat zone segmentation of fibrous textures is without interest for construction descriptors which characterize the morphological properties of the fiber network.

Hence we offer to use the Local Radon Transform (LRT), as discussed in [10], to segment the individual fibers. The LRT uses an orientated Gaussian derivative kernel, which is rotated at different angles and adapts via a maximization procedure to the various directions of the texture (see Fig. 3). Now, the computation of connected flat zones (two neighboring points belong to the same zone if they have the same value of dominant orientation) produces a segmentation of the network in J linear segments which roughly represent each fiber, but which do not cover the whole support space, i.e., $\cup_{j=1}^{J} F_j[f] \neq E$; in addition, in the crossing points of fibers one of the orientations is arbitrary favored over the others. We have solved this last drawback by considering separately the connected components of each orientation and by reconnecting them by a small morphological closing. Once the segmentation of individual fibers is available, each fiber $F_j[f]$ can be described by its length l_j (computed as its geodesic diameter) and by its orientation θ_j. Using these two parameters, we propose to characterize a fibrous texture f by a new statistical matrix named Orientation Length Zone Matrix (OLZM), $\mathcal{OL}_f(\theta_n, l_m)$, which yields the number of "fibers" of orientation θ_n and length l_m. The number of rows of OLZM depends on the degree of discretization of the orientation space (which is selected in the computation of the LRT) and the number of columns equal to the longest fiber of the texture.



(a) (b) (c)

Fig. 3. The original microtubule network (a); an enhanced network image (with LRT) giving at each pixel $\mathbf{x}$ the value of the image processed with kernel orientated at angle θ_i which produces the maximal intensity (b); an orientation map giving, at each $\mathbf{x}$, the corresponding maximal response θ_i represented by a different color label (c).

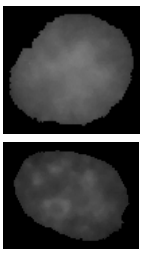
To deal with the problem of rotation invariance, we offer the following alternative solutions. If the texture f is a segmented object, for instance a cell such as in our case study, the orientation of each zone is given with respect to the coordinate system associated to the main axis of the object (computed by PCA). If the texture f is not bounded in the image, the main axis can be replaced by the average of orientations of all the zones which compose the texture.

5. APPLICATION TO CELL CLASSIFICATION IN CELL DIVISION ASSAYS

The development of the present texture descriptors was motivated by an application in cell-based assays for phenotypic screening, which consists in the use of multi-parametric and high resolution imaging techniques to characterize and select innovative compounds and/or protein targets involved in cell division. More precisely, the aim is to build a cell phase classifier by the analysis of the DNA structure (using a marker of the nuclear chromatin) and the organization of microtubule network (using a marker of the cytoskeleton).
We dispose of a set composed by 317 cells which are labelled by experts: as interphase or phase of the mitosis and approximately 50 other labels that provides relevant information for phase classification. To elaborate classification models, we realize *One Class Classifiers* for each class with logistic regression and validation by *Leave One Out* (k-fold validation with k equal to the number of instances).

Among these labels, we focus here on:

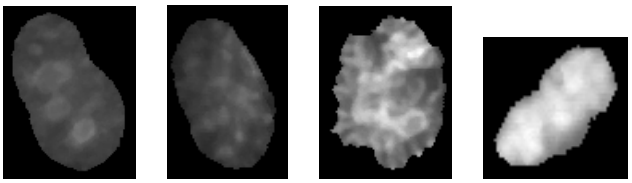
1. *Texture Homogeneity* of the DNA, which contains two classes *Homogeneous* and *Heterogeneous*. Fig. 4 shows that MGLSZM provides comparable result to the best original GLSZM. This property can be observed too on Fig. 5 for radial distribution. These results demonstrate the power and usefulness of such multiple gray level version.



Matrix	Prediction
$\mathcal{GS}_f^8(s_n, g_m)$	$89,3$
$\mathcal{GS}_f^{16}(s_n, g_m)$	$91,1$
$\mathcal{GS}_f^{32}(s_n, g_m)$	$92,7$
$\mathcal{GS}_f^{64}(s_n, g_m)$	$91,6$
$\widetilde{\mathcal{GS}}_f(s_n, g_m)$	$92,7$

Fig. 4. Examples of homogeneous texture (top left) vs. heterogeneous texture (bottom left) nuclei; and results of homogeneity classification using second-order moments of GLZSM.

2. *Masses Texture* of the DNA, which is the "distribution" of DNA and contains four classes (*Beads*, *Slightly Condensed*, *Condensed* and *Highly Condensed*). Fig. 5 shows that (M)GLDZM provides a better description of radial distribution than MGLSZM.
However, there is an exception for class "Beads", which is composed of texture without radial repartition (so random or homogeneous texture). Thus MGLSZM is well adapted and provides better results.

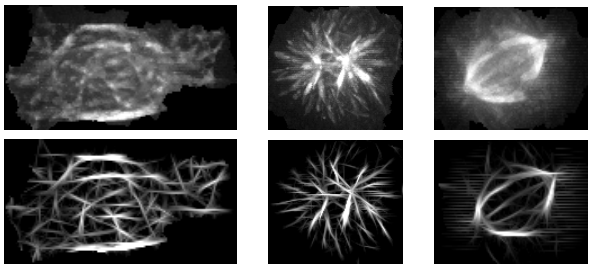


(a) (b) (c) (d)

Matrix	Beads	Slightly	Condensed	Highly
$\widetilde{\mathcal{GS}}_f(s_n, g_m)$	$91,3$	$88,1$	$89,75$	$92,2$
$\mathcal{GD}_f^8(d_n, g_m)$	$85,7$	$85,8$	$86,1$	$89,1$
$\mathcal{GD}_f^{16}(d_n, g_m)$	$88,1$	$87,6$	$90,75$	$95,3$
$\mathcal{GD}_f^{32}(d_n, g_m)$	$86,1$	89	92	93
$\widetilde{\mathcal{GD}}_f(d_n, g_m)$	$89,75$	89	$91,9$	$95,05$

Fig. 5. Examples of nuclei with different texture (associated here to the condensation of the chromatin): beads (a), slightly (b), condensed (c), highly (d); and results of classification.

3. *Network Organization* of microtubules, which contains three classes (*Well Organized*, *ReOrganized* and *Other*). Fig. 6 shows really satisfying of OLZM. However, these results are slightly lower than previously due to the organization similarity among texture classes.



(a) (b) (c)

Matrix	Well Organized	ReOrganized	Other
$\widetilde{\mathcal{OL}}_f(d_n, g_m)$	$86,1$	$77,8$	$84,9$

Fig. 6. Examples of microtubule network normal/enhanced with different organization: Well Organized (a), ReOrganized (b), Other (c); and results of classification.

6. CONCLUSION AND PERSPECTIVES

In this paper, the problem of cell phase mitosis classification is addressed. More precisely, we focus on specific and relevant parts of cell characterization particularly difficult. To figure these problems out, we design new advanced statistical matrices based on Gray Level Size Zone Matrix. First a multiple gray level version which uses more information about the texture, and next two new versions which use radial distribution, length and orientation of flat zones. These matrices showed their power and efficiency and can be applied to other problems of texture characterization.

Acknowledgements. This work is part of (2007-2010) RAMIS project funded by the General Directorate for Competitiveness, Industry and Services of the French Ministry for the Economy, Industry and Employment. The authors thank Chantal Etievant, Dephine Reberioux, Vanessa Tillement and Michel Wright from Pierre Fabre for providing the annotated cell images, and Ronan Danno from ADCIS company for the annotation software.

7. REFERENCES

[1] Jianguo Zhang and Tieniu Tan, "Brief review of invariant texture analysis methods," *Pattern Recognition*, vol. 35, no. 3, pp. 735–747, March 2002.

[2] Robert M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 3, no. 6, pp. 610–621, 1973.

[3] M. M. Galloway, "Texture analysis using grey level run lengths," *Computer Graphics Image Processing*, vol. 4, pp. 172–179, July 1975.

[4] A. Chu, C. M. Sehgal, and J. F. Greenleaf, "Use of gray value distribution of run lengths for texture analysis," *Pattern Recognition Letters*, vol. 11, no. 6, pp. 415–419, 1990.

[5] Michael Unser, "Texture classification and segmentation using wavelets frames," *IEEE Transactions on Image Processing*, vol. 4, no. 11, pp. 1549–1560, November 1995.

[6] Guillaume Thibault, Bernard Fertil, Claire Navarro, Sandrine Pereira, Pierre Cau, Nicolas Levy, Jean Sequeira, and Jean-Luc Mari, "Texture indexes and gray level size zone matrix. application to cell nuclei classification," in *Pattern Recognition and Information Processing (PRIP)*, Minsk, Belarus, May 2009, pp. 140–145.

[7] Robert M. Haralick, "Statistical and structural approaches to texture," in *Proceedings of the IEEE*, May 1979, vol. 67, pp. 786–804.

[8] Fernand Meyer, "The levelings," in *Mathematical Morphology and its Applications to Image and Signal Processing*, Heijmans and Roerdink Eds, Eds., Kluwer, 1998, pp. 199–206.

[9] L. Decker, D. Jeulin, and I. Tovena, "3d morphological analysis of the connectivity of a porous medium," *Acta Stereologica*, vol. 17, no. 1, pp. 107–112, 1998.

[10] M. Krause, R. M. Alles, B. Burgeth, and J. Weickert, "Retinal vessel detection via second derivative of local radon transform," Tech. Rep. 212, Department of Mathematics, Saarland University, June 2008.