ARITHMETIC METHODS
IN MATHEMATICAL PHYSICS AND BIOLOGY
BANACH CENTER PUBLICATIONS, VOLUME 109
INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES
WARSZAWA 2016

FRACTAL ANALYSIS IN DIGITAL HISTOLOGY

HELMUT AHAMMER, MICHAEL MAYRHOFER-REINHARTSHUBER
and PHILIPP KAINZ

Institute of Biophysics, Medical University of Graz
Harrachgasse 21/IV, 8010 Graz, Austria
E-mail: helmut.ahammer@medunigraz.at,
michael.mayrhofer-reinhartshuber@stud.medunigraz.at, philipp.kainz@stud.medunigraz.at

Abstract. Digital images are increasingly used in medicine, especially in digital pathology. Histological techniques are well established and modern digitalization systems yield high resolution digital images. Evaluation of these images is regularly accomplished by subjective inspection, but objective or numerical methods are still rare. Biological patterns and textures are mathematically hard to measure or to simulate and therefore, fractal methods and non-Euclidean geometry are very suitable to solve these tasks.

Digital images inherently have many advantages and fractal measures can be calculated by various methods, but care has to be taken in order to gain reliable and robust results. Digital images are discrete representations of specimen and can be noisy. Furthermore, an image is not always totally filled by the specimen and consequently background pixels must be considered, too. If appropriately considered, these issues do not decline the power of fractal analyses.

1. Introduction. Histology is the science of biological tissue and is therefore a subdomain of medicine and biology. Particularly, it is a subdomain of anatomy and pathology. First, a diagnostically conclusive part of a tissue is cut out and fixed by detergents such as formaldehyde. The water content of the tissue must be reduced and minimized by rinsing in alcohol or impregnation with paraffin. Paraffin also fixes the tissue. Another embedding agent can be acrylate resin or alternatively shock freezing. Tissue blocks conserved in this way are morphologically and temporally stable and can be stored for several years. For diagnostic purposes, very thin slices of about 2–10 μm are cut off by using a microtome. These very fragile slices are carefully placed on a slide and stained with special dyes. One of the most used stainings is the H&E (Haematoxylin and Eosin) stain. Haematoxylin

2010 *Mathematics Subject Classification*: Primary 28A80; Secondary 92C50.
Key words and phrases: Fractal Analysis, Fractal Dimension, Digital Images, Limited Resolution, Image Background, Image Noise, Histology, Digital Pathology.
The paper is in final form and no version of it will be published elsewhere.

DOI: 10.4064/bc109-0-2

stains the cell cytoplasm or collagen fibers pink and Eosin stains cell nuclei typically cyan blue. Staining of specimen is an essential step because otherwise contrasts would be too low by inspection through a light microscope. Diagnosis is commonly carried out subjectively but is prone to inter as well as intra observer variance. Recently, digital techniques have been developed and new scientific disciplines such as digital pathology or computer-aided diagnostics have been emerged [AJHD, GBCMRY, GKGM, HLBD, W, YG].

Digital images of tissue specimen can be taken with high spatial resolution using microscopes or modern whole slide scanners and diagnosis can be carried out by using high resolution monitors. These new techniques open a wide range of possibilities for mathematical preprocessing, segmentation, or classification applications using simple algebra, convolution or partial differential equations including optimization algorithms. However, these rather objective analyses are rarely used in medicine because the gap between mathematical algorithms and clinical practice is often too high.

In order to circumvent this problem it is necessary to develop mathematical algorithms which are reliable, transparent and easily comprehensible. One prominent theory, which fulfills these requirements, is the theory of fractals and non-Euclidean geometry developed by Mandelbrot [M].

2. Image acquisition and image preprocessing. Digital images of medical samples in pathology and histology are regularly taken by microscopes or modern whole slide scanners [GBCMRY, HLBD, YG]. Microscopes achieve high optical magnifications featuring high optical quality. High performance digital cameras are widely available and therefore, high resolution digital images can be taken without much effort. A drawback may be the limited field of view, particularly for high magnifications. Sample images for some magnifications and corresponding field of views can be seen in Figure 1, e.g. a magnification of 40× leads to a field of view of only about 0.5 mm.

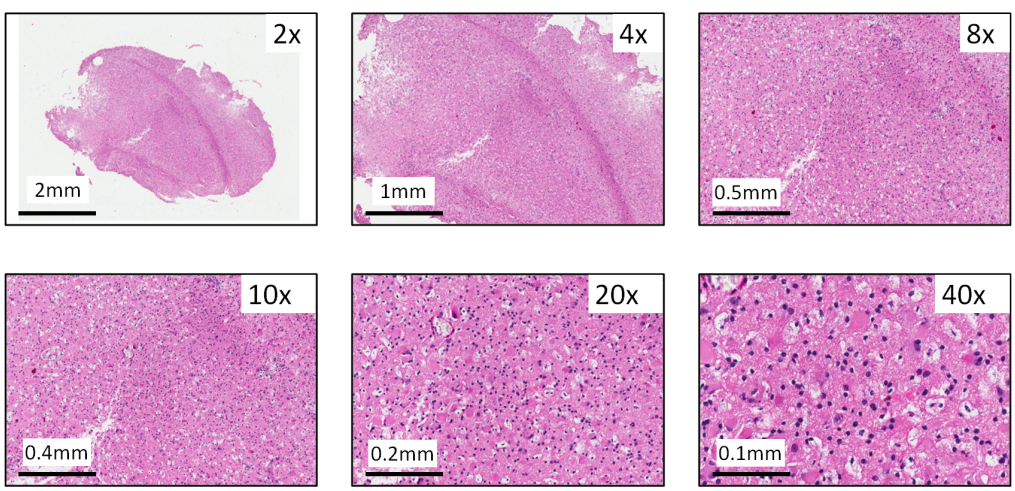

Fig. 1. Magnification and field of view. Histological images of the brain. Increasing magnification yields a decreasing field of view

Thus, microscopes are well suited to take high resolution images of small specimen such as isolated cells in the micrometer scale. Otherwise, tissue comprising several cells in the millimeter or even centimeter scale can be scanned with whole slide scanners. Magnifications are not as high as compared to microscopes, because immersion oil cannot be used. Nevertheless, the resolution is high enough to visualize cells and cell compartments such as the cell nucleus. For both methods digital resolution can be quite high as hardware is continuously improving. Therefore, whole slide scanner images can reach a pixel resolution up to 50.000×50.000 pixels and more. Although hardware equipment enables taking high quality digital images it is not granted that a quantitative mathematical analysis can be performed without any issues. Images are often taken by the operator in order to fulfill subjective needs or requirements in order to make a diagnosis and nothing more. Usually, quantitative algorithms have higher demands on image quality and especially on variations from image to image. Variations of brightness, color, contrast and background should be as low as possible. These variations are hard to bypass and therefore, sophisticated image preprocessing steps prior to the actual analysis are frequently unavoidable [GW, P].

First, white color must be checked or adjusted in order to have a homogeneous and clear defined background as well as clear and crisp colors without tint. Sometimes images must be cropped because the specimen is smaller than the field of view or other objects inside the image must be eliminated. Most often images have to be segmented to gain just image parts or objects of interest (e.g. the nucleus of a cell without stroma and other cell compartments). Segmentation can be carried out in the simplest form by applying an intensity threshold using the histogram of an image. There exist several algorithms incorporating an error functional for optimal thresholding. Simple methods are Otsus method, Energy or Entropy concepts and more complicated algorithms are total variation or the use of neuronal networks [P]. Figure 2 shows a sample histogram of a histological

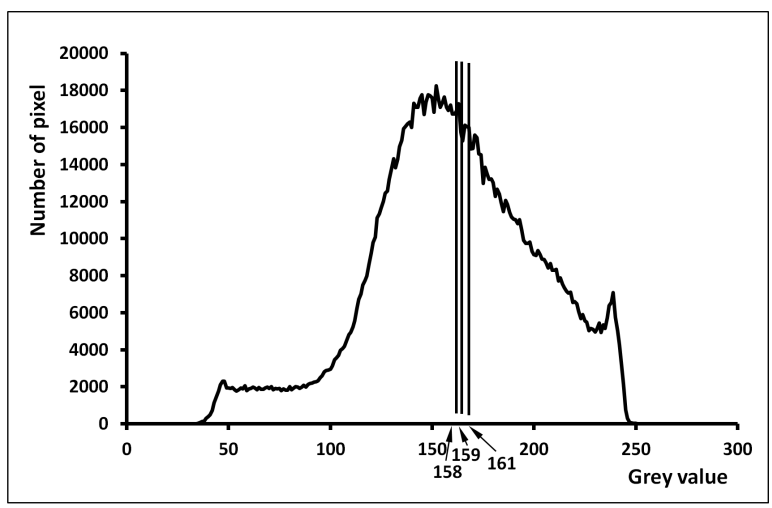

Fig. 2. Histogram of a histological image. The image with 40× magnification was color to grey value converted and the threshold values for Otsu's method and the Energy and Entropy values were calculated. These three threshold values are depicted as vertical lines

image (actually a color to grey converted version of the image with 40× magnification in Figure 1) and the calculated Otsu's, Energy and Entropy threshold values.

Image stacks of a 3D representation of tissue are a special hassle because they must be registered, too. Although registering can be easily applied to magnetic resonance (MR) or computed tomography (CT) images, it is very hard to achieve for histological images [D]. Slices can be shifted, rotated or even deformed and only very sophisticated algorithms such as elastic registering by optical flow can eliminate these artifacts. If of interest, it is necessary to segment edges by using convolution kernels or filter techniques. The result of the segmentation step can be in form of grey values or of a binary representation.

It must be noted that image segmentation is almost always an unavoidable prerequisite and the quality of image segmentation inevitably determines the quality of subsequent quantitative analyses. This is a very important fact, which must not be trivialized.

3. Fractal analysis of digital images. The concept of fractals is tightly connected to nonlinear chaotic dynamics [F]. Chaotic systems are deterministic systems with a very high sensitivity to initial conditions leading to attractors in a parameter phase space. Chaotic attractors show self similarity and cannot be described satisfactorily by Euclidean geometry. So, they are fractals and represent a geometrical footprint of the underlying dynamical system. Weather, lasers, heart beat dynamics and many other natural dynamical systems are directly investigable in the temporal domain as well as the phase space spatial domain. In case of histological or histopathological images only the spatial domain is investigable, but it seems reasonable to suppose that these images can be interpreted as visual time stamps of a dynamical system producing the objects inside an image. However, it will never be possible to gain or measure the dynamical system directly, as this dynamical process can have taken years or even decades of life.

Therefore, interpreting static digital images as fractals is reasonable, but it must be stated that care has to be taken, in order to not abuse or misinterpret fractal concepts. The following points have to be reviewed critically before applying any algorithm for calculating e.g. fractal dimensions.

(i) *Digital images have only a limited range of scales.* Simply the image width or height is the largest scale and the size of a single pixel is the smallest scale available. Assuming e.g. a rather high resolution image comprising 10.000×10.000 pixels we get only four orders of scales using decimal powers (the fifth order contains only a single pixel and cannot be interpreted seriously as an image). Therefore, fractal concepts cannot be applied with their strict mathematical definitions, mainly because the limit of infinitesimal scales cannot be calculated. Therefore, objects in a digital image can only be interpreted as being statistically fractal and calculations lead to rather estimations than exact determinations. In a broader sense these restrictions are not really surprising, because often nature can only be described satisfactorily in a statistical way. In physics, very powerful theories such as thermodynamics as well as quantum mechanics are statistical theories and only probabilities can be calculated. Nevertheless, these theories have proven to be the best way in order to describe the corresponding physical conditions.

(ii) *Decision about which image compartment should be treated as a fractal.* It is necessary to assign the right topological dimension of the object or objects under investigation because e.g. the fractal dimension is always equal or greater than the topological dimension. Histological images can appear in various manifestations and a very important parameter is the effective field of view that is covered by the digital image. The extent of tissue under investigation can be larger than the effective field of view. Such an image is filled with tissue only and does therefore not contain any tissue borders or background (area without any tissue). In this case, the structure of cells or cell nuclei and their spatial relationships can be very well interpreted as being fractal. Particularly, the grey level surface of such images can be investigated and therefore the topological dimension should be assigned to be 2. This evaluation uses every part of the image (tissue) and therefore, is not specific. If a specific evaluation is necessary, e.g. if only the spatial relationship of cells or cell nuclei is of interest, it is necessary to segment the image prior to fractal analyses. Image segmentation can be performed by using any method that works well enough and is not restricted to special methods for fractal analyses. Segmented images are usually binary images composed of only two grey levels, in particular black and white. In this case it is however necessary to assign a topological dimension. The topological dimension is 2 for larger object areas (cells or nuclei) but it can also be zero in case for very low magnifications when cells or cell nuclei rather appear as points than as areas. The segmented binary image can be used to investigate another characteristic of the tissue, namely the border of cells or cell nuclei. It is very easy to gain these borders in binary images, e.g. via morphological erosion or dilation followed by image subtraction. Nevertheless, it is possible to extract borders directly from a grey value image by using several image processing techniques such as convolution in the spatial or frequency domain. Conclusively, borders appear as lines in an image and must be assigned to a topological dimension of 1.

(iii) *Objects with background.* The aforementioned aspects and decisions are harder to gain when the objects of interest do not fill the total area of an image. Tissue may not fill the field of view or the object itself is a single cell or e.g. a cell cluster, surrounded by non-tissue medium. This background is usually black for grey level images and white for color images. Without consideration most algorithms will calculate inaccurate values since background can be flat with a constant grey value or noisy with slightly varying grey values. Visually, the difference between these two cases may not be noticeable because slight grey value changes, e.g. in the range of 1 to 10, are not differentiable by our human visual system. For a defined brightness only about 16 grey levels can be distinguished visually. On the other hand background can decrease or increase values of the fractal dimension, because a really flat and constant area is ordered and a variable area introduces some noise, in most cases random noise. Thus, it is mandatory to circumvent this uncertainty by identifying background pixels by image segmentation and by assigning a predefined grey value, usually zero. Some pixels inside the object may be zero too, but this can be avoided by increasing all pixel grey values by one before segmentation.

(iv) *Image noise.* Every digital image has been taken by a technical device and the brightness reflects a physical process or interaction with the specimen. The interaction involved and the saving of data is in essence a physical measurement and therefore, noise

is unavoidable. Noise in digital images can have various reasons; it can be random noise, Brownian noise, shot noise or any other type of noise. In case of random noise the grey value fluctuates additionally in a random manner and therefore the fractal analysis will be affected, mainly by increasing fractal dimensions, which can be seen in Figure 3 (first column) [AD].

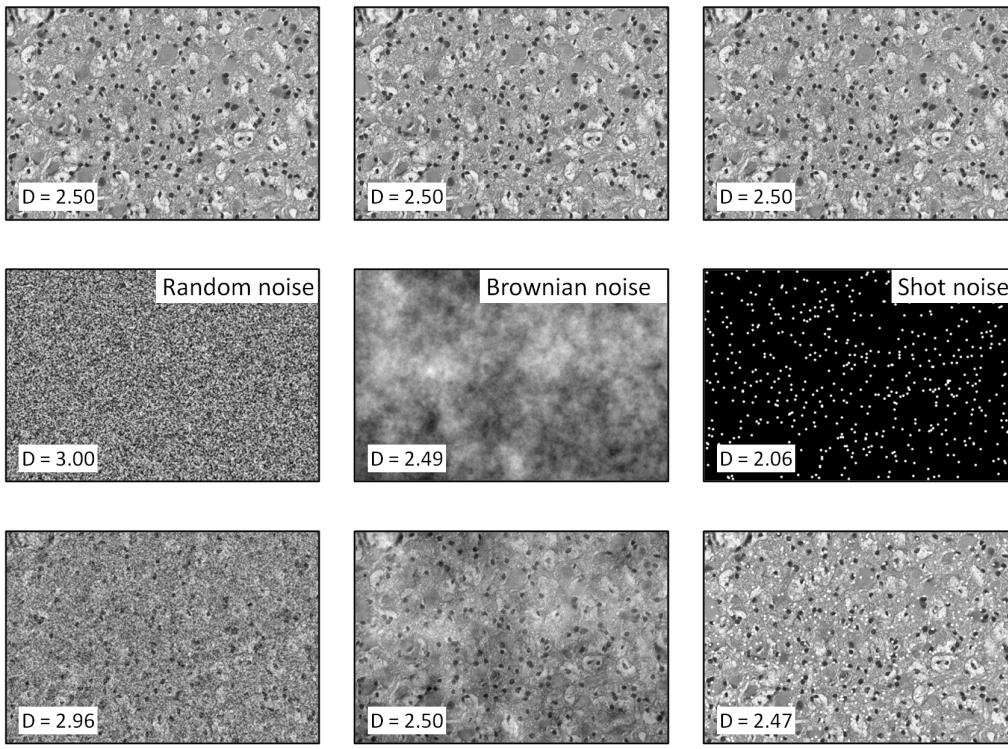

Fig. 3. Influence of image noise. The first row shows a histological image with fractal dimension $D = 2.5$. The second row shows images of random noise, Brownian noise and shot noise with their corresponding fractal dimensions. The third row shows a columnwise combination of the images in the first and second row and the resulting fractal dimensions

In this example the fractal dimension of the Brownian noise with $D = 2.49$ (second column, second row) is very near to the value of the tissue itself and therefore, the resulting value is practically unchanged (second column, third row). With the low fractal dimension value of 2.06 for the shot noise (third column, second row) the resulting fractal dimension is markedly decreased (third column, third row).

Hence, image preprocessing by noise filtering should always be considered and very often it is necessary to filter noise by using low-pass filtering, Median filtering or more sophisticated methods such as total variation (TV) methods.

After critically following (i)–(iv) fractal dimensions of digital images can be calculated, but at this point it must be stressed that subsequent calculations strongly depend on these preprocessing considerations and preprocessing steps actually performed.

4. Fractal analysis of binary images. Binary images have a lower information content compared to grey level images and the analysis can be easily implemented by counting procedures. The most prominent and effective methods are the well known Box counting method, the Minkowski method and the Correlation method. Box counting uses several box side lengths and the number of boxes overlapping the object is counted for each side length. Box counting can be equivalently performed by using an image pyramid, which is an ensemble of images with decreasing resolution (image side length). The number of object pixels is counted. For rectangular images the calculated values are strictly identical, with the advantage of higher computational speeds and less coding efforts. Another well known method is the Minkowski method, which uses morphological dilation (Minkowski dilation) in order to gain dilated object areas [KRR]. Finally, the correlation method counts the number of neighboring pixels for various radii. All these methods share a common aspect: they calculate values (numbers) for various lengths (scales). A double logarithmic plot is constructed and a linear regression gives a slope which can be used to calculate an estimate of the fractal dimension.

5. Fractal analysis of grey level images. Grey value images show the full amount of image information and methods for calculating the fractal dimension mainly focus on the grey value surface. This surface is embedded in three dimensions (two spatial dimensions and one grey value dimension). The Differential Box counting method is an adaption of the Box counting method, which fits boxes with various side lengths under the surface [SC]. Accordingly, the Blanket dimension is an adaption of the Minkowski dimension and calculates blankets of increased as well as decreased surfaces [DQRTZ]. Another method uses 2D Fourier transformation, particularly FFT in order to calculate the power spectrum of the grey values and the slope of the falling spectrum gives an estimate of the fractal dimension [DQRTZ].

The list of mentioned methods calculating the fractal dimension is far from complete, but here we would like to focus on some common aspects which should be considered or at least mentioned.

If areas of objects increase, the estimate of the fractal dimension will increase, too. This can be explained by the space filling properties of fractals, which is measured by the fractal dimension. Increased areas can be an inherent property of one object compared to another one, but for a given object actual settings for image preprocessing and image segmentation can produce a slightly smaller or a slightly larger area for a single object. This fact can also be noticed for image borders. Some algorithms produce very small border lines on a single pixel level, but other algorithms produce broader borders. Therefore, e.g. increased fractal dimensions will be estimated when the segmented border is thicker.

The background of an image may strongly influence the calculations of fractal dimensions. This can be easily simulated by adding a border of background pixels to an image and calculating the Box counting dimension (see Figure 4). Additional background pixels do not hit the object and therefore, the relative count will decrease or sometimes increase with increasing border.

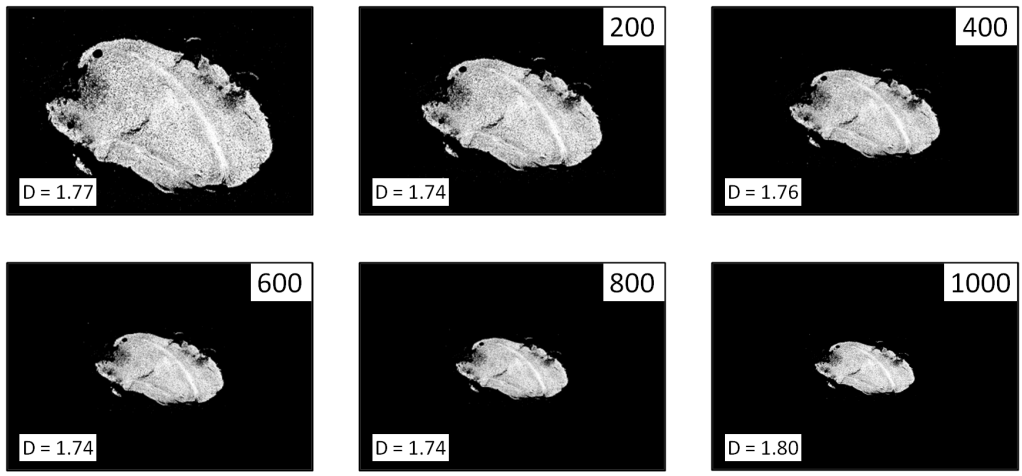

Fig. 4. Influence of background border. Segmented binary sample images (original image in Figure 1, 2× magnification, 1568 × 1058 pixel) with increasing number of border pixels. Fractal dimensions were calculated using the Box counting method

On the other hand, the Minkowski method for binary images does not depend on background pixels as it uses dilations of the object pixels, which does not per se depend on background pixels. An additional background border would not change the result. Actual fractal dimension values for the images in Figure 4 were identically $D = 1.79$. Nevertheless, this is true only for compact objects without background pixels inside the objects such as single cells or cell spheroids. Considering images of tissue, background pixels inside the tissue can be caused by areas of missing tissue at all, e.g. pulled out tissue by mistake during the histological slicing procedure. In these cases a correction for background pixels is necessary, but far from easy to accomplish. The most feasible way would be segmenting background areas by image processing and using the result as a mask for further calculations. For grey level images it would be necessary to omit calculations for background pixels at all. Fourier transformation does not inherently sum up values that are zero and therefore a black background outside or inside the objects does not count.

Image noise potentially increases the heterogeneity of the spatial grey level distribution and may increase the absolute values of fractal dimension estimates. Figure 5 shows a simulation of this noise dependency. Outgoing from a grey value image without noise and a signal to noise ration $SNR = 1$, random noise was added until $SNR = 0$. Accordingly, the values for the fractal dimensions increased from $D = 2.5$ to $D = 3.0$.

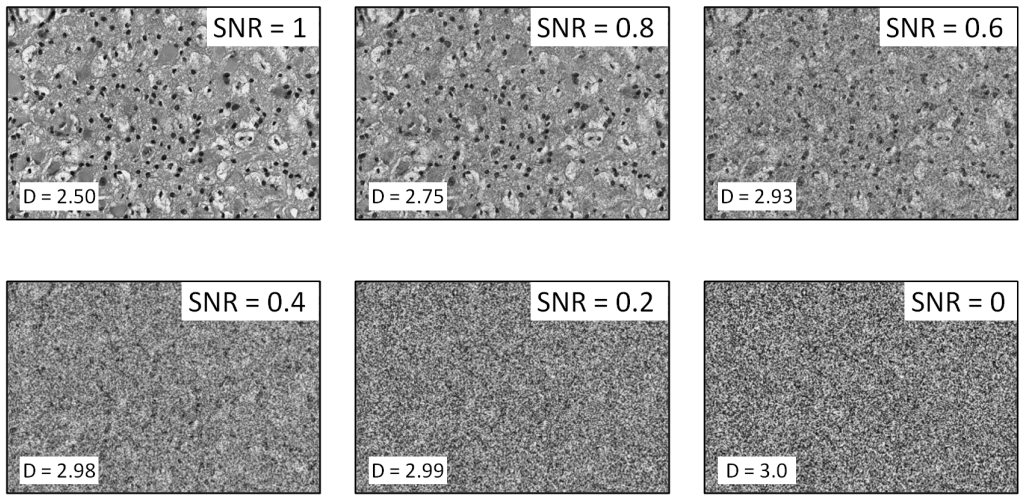

Fig. 5. Influence of decreasing signal to noise ratios from $SNR = 1$ to $SNR = 0$. With increasing noise, the values for the fractal dimensions increase correspondingly

Image denoising by image preprocessing should be applied before performing any calculations, but the absolute influence of noise may still remain unclear. A possible solution could be an empirical study of simulating additional predefined noise levels with a noise free sample image. In most cases this noise free sample image must be artificially constructed using a simulation model, a manual drawing or a segmented real sample image.

6. Conclusion. In recent years, fractal analyses of histological digital images have proven to be reliable in order to gain knowledge about the biological status of a specimen. In digital pathology grading of distinct tissue samples is possible by evaluating single cells, cell nuclei or the whole tissue at all. Nevertheless, digitized images introduce some limitations that must be considered before applying theoretical fractal concepts. Main issues are the discrete image space, the background and the image noise. If well considered, estimations of fractal measures can be gained, which are reliable and useful for future applications in histology and digital pathology.

References

[AD] H. Ahammer, T. T. J. DeVaney, *The influence of noise on the generalized dimensions*, Chaos Solitons Fractals 26 (2005), 707–717.

[AJHD] S. Al-Janabi, A. Huisman, P. J. Van Diest, *Digital pathology: current status and future perspectives*, Histopathology 61 (2012), 1–9.

[D] T. M. Deserno (ed.), *Biomedical Image Processing*, Springer, Heidelberg 2011.

[DQRTZ] B. Dubuc, J. F. Quiniou, C. Roques-Carmes, C. Tricot, S. W. Zucker, *Evaluating the fractal dimension of profiles*, Phys. Rev. A 39 (1989), 1500–1512.

[F] K. Falconer, *Fractal Geometry: Mathematical Foundations and Applications*, 3rd ed., Wiley, Chichester 2014.

[GKGM] K. Glatz-Krieger, D. Glatz, M. J. Mihatsch, *Virtual slides: high-quality demand, physical limitations, and affordability*, Human Pathology 34 (2003), 968–974.

[GW] R. C. Gonzalez, R. E. Woods, *Digital Image Processing*, 3rd ed., Prentice Hall, Upper Saddle River 2007.

[GBCMRY] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, B. Yener, *Histopathological image analysis: a review*, IEEE Rev. Biomed. Eng. 2 (2009), 147–171.

[HLBD] A. Huisman, A. Looijen, S. M. van den Brink, P. J. van Diest, *Creation of a fully digital pathology slide archive by high-volume tissue slide scanning*, Human Pathology 41 (2010), 751–757.

[KRR] S. Kanmani, C. Babu Rao, B. Raj, *On the computation of the Minkowski dimension using morphological operations*, Journal of Microscopy 170 (1993), 81–85.

[M] B. Mandelbrot, *How long is the coast of Britain? Statistical self-similarity and fractional dimension*, Science 156 (1967), 636–638.

[P] W. K. Pratt, *Digital Image Processing: PIKS Scientific Inside*, 4th ed., Wiley-Interscience, Hoboken 2007.

[SC] N. Sarkar, B. B. Chaudhuri, *An efficient differential box-counting approach to compute fractal dimension of image*, IEEE Trans. Systems Man Cybernet. 24 (1994), 115–120.

[W] R. S. Weinstein, *Innovations in medical imaging and virtual microscopy*, Human Pathology 36 (2005), 317–319.

[YG] Y. Yagi, J. R. Gilbertson, *A relationship between slide quality and image quality in whole slide imaging (WSI)*, Diagnostic Pathology 3 (2008), S1–S12.