Image Processing in Biology Based on the Fractal Analysis

István Sztojánov¹, Daniela Alexandra Crișan², Cătălina Popescu Mina³ and Vasilică Voinea¹

¹Politehnica University of Bucharest
²Romanian-American University, Bucharest
³University of Bucharest
Romania

1. Introduction

The state of the art of classification based on the fractal analysis with applications in biology will be presented. Fractal features describe closely the properties of natural forms. For this reason, the interest in this new mathematical field, fractal geometry, grows quickly. New techniques of fractal analysis are developed and these techniques prove their utility in real systems in various fields such as informatics, economics, engineering, medical screening and biology.

This chapter discusses problems of classification based on the fractal theory with applications in biology. Here are introduced the necessary notions for the defining of the fractals and their quantitative evaluation and an algorithm for fractal dimension computation based on biofractal contours processing is also presented. Concretely, there were extracted window-images from the interested area; the contours of the window-images were processed with the “box-counting” algorithm in order to establish the fractal dimensions for the analysed sections. The algorithm based on the “box-counting” method offers two major advantages: it is easy to implement in case of using a computer and can be applied for images no matter how complex.

The first application is dedicated to the analysis of the particularities of some species from Gentianaceae family, with the purpose of establishing their affiliation to the Gentiana genus, knowing the fact that, up to the present, there have been used only evaluations based on the distinctive morphological characteristics. Concretely, there were extracted window-images from the rind and the central cylinder of the root and stem and also from the mesophyll and leaf nervure/rib and those areas where analysed by using fractal techniques. We consider that the acquiring of samples from more sections of a species (the studied one) for the statistical processing of the data will lead for the first time in botany to very precise characterizations of that species.

The second applications use the same procedure in order to classify mammary tumours: benign or malign. The fractal dimension of the FAR (Focused Attention Region) is computed, by using the same box-counting algorithm. Depending on the size of the fractal dimension, a classification can be made: over 30 lesions cases with known diagnostic were...
tested and the results show that a threshold of 1.4 value can be used: over 90% malign cases have fractal dimensions above 1.4, while over 80% benign cases have fractal dimensions below 1.4.

The results are encouraging for the development of fractal techniques in biology domains.

2. Fractal analysis in biology

The fractal analyses proved their utility in biology and the necessary notions for the defining and the quantitative evaluation of the fractals will be presented; an algorithm for fractal dimension computation based on biofractal contours processing will be shown.

2.1 Biofractal and fractal dimension

Fractals are objects with irregular, auto-similar features, with details that can be noticed at any scale of representation. Biofractals are the fractal textures/contours in biology (tissues, neurons, leaves, etc.). The similarity between fractals and the natural objects suggests that fractal properties, such as fractal dimension, may be used as a classifier in biology.

In order to understand the surrounding world, the natural sciences have progressed by focusing on the simplest forms of representation, in accordance with the principle: simplicity explains complexity. Sometimes the researches turned away from the direct study of nature, of the details of reality and limited themselves to studies based on general, approximate and linear expressions.

In the ‘60s, a mathematician, Mandelbrot proposes to study the complex irregular forms in nature that he names fractals and founds the bases of fractal (non-Euclidian) geometry. The appearance of fractal geometry marks the return of the scientific knowledge to the real world. It was rather easy to observe that the forms of rivers, mountains, the Earth in its details are of fractal type. Important examples of fractals in botany – that we shall name fitofractals - are the leaves of a tree, the structure of the tissue from a plant’s stem or root section, the forms and contours of the cells etc.

With the aid of fractal geometry, the growth and ramification models from the plants world can be explained and reproduced (fig. 1.) by using strings of ordered characters and simple operations of translation and rotation (Lindenmayer & Prusinkiewicz, 1996).

Fig. 1. The construction of an L-system that can be applied to ramification types of plants

An important feature of the fractal objects is the dependence between their dimension and the used measure unit (fig.2.). By choosing a finer measure unit, an irregular contour can be better approximated, with finer accuracy, and this is so because a finer measure unit better shows the objects details.
Richardson noticed this fact for the first time. Desirous to know the length of the borderline between Spain and Portugal, he consulted the documents from the archives of both countries. In the Spanish encyclopaedia it was written that the borderline had a length of 987 km, while in the Portuguese encyclopaedia the length was approximated to 1214 km. The explanation of the strange phenomenon was that two different measure units were used: the smaller unit used by the Portuguese could go over more border details, and so they obtained a finer measurement.

This dependence from the used scale measure makes the fractal objects difficult to measure in the classic (Euclidian) geometry context.

The Euclidian dimension $D$ is given by the number of coordinates needed to define any of the points of the object, or, more exactly, the dimension of the Euclidian space into which the analysed object can be submerged: the line in a plane, the cone in a three dimensional space.

The topological dimension $T$ is defined by the local properties of the analysed object and corresponds to the concept that the dimension of a point is 0, the dimension of a line, thin curve is 1, the surfaces have the dimension equal with 2, volumes with 3 and so on, without taking into account any bigger dimension of the Euclidian space into which these forms were submerged.

With the apparition of the fractals, the characterization of a form by using its topological dimension (which is a whole number) proves its insufficiencies. That is why the notion – fractal dimension $D_f$ (real number) was introduced.

The German mathematician Felix Hausdorff defines a new concept for the topological spaces, in this way suggesting that the fractal dimension is proportional with the minimum number of spheres, of a given radius, needed for covering the measured object. To facilitate the computer work, the coverage is made with cubes instead of spheres.

Thus, for covering a curve of unity of length 1, $N(s)=1/s$ cubes of side $s$ are needed, for covering a unity area surface there are needed $N(s)=1/s^2$ cubes of side $s$ and finally, to cover a unity volume cube $N(s)=1/s^3$ cubes of side $s$ are needed (fig 3).

By induction the relation below is verified as follows:

$$N(s) \sim 1/s^D,$$

where:

- $N(s)$ is the number of cubes of side $s$;
- $s$ is the scale coefficient or the length of the coverage of the cube’s side;
- $D_f$ is the Hausdorff’s coverage dimension of the object.
By applying logarithm to the relation above, we can deduce $D_f$:

$$D_f = \frac{\log(N(s))}{\log(1/s)}$$  \hspace{1cm} (1)

The fractal dimension $D_f$ is however difficult to calculate.

2.2 Algorithm for fractal dimension computation

There can be used many algorithms to calculate the fractal dimension and one of the easiest to implement is the box-counting algorithm. The algorithm based on the "box-counting" method offers two major advantages: it is easy to implement in case of using a computer and can be applied for images no matter how complex. The "box-counting" fractal dimension, derived from the Hausdorff coverage dimension is given by the following approximation:

$$D \approx \frac{\log(N(s))}{\log(1/s)}$$  \hspace{1cm} (2)

It is expected, that for a smaller $s$, the above approximation should be better,

$$D = \lim_{s \rightarrow 0} \log \frac{N(s)}{\log(1/s)}$$  \hspace{1cm} (3)

If this limit exists, it is called the "box-counting" dimension of the measured object. Usually, this limit converges very slowly, that is why an alternative solution is used. Since the expression:

$$\log(N(s)) = D \cdot \log \left( \frac{1}{s} \right)$$  \hspace{1cm} (4)

is the equation of a straight line of slope $D$, the "log-log" curve described by the points of the $(\log(N(s)), \log(1/s))$ coordinates is plotted. Through linear regression (least squares method) the slope of the line that approximates the points' distribution is determined; this is the wanted fractal dimension.
Thus, the regression line has the form:

$$Y = a \cdot X + b$$

(5)

and the line slope (the value of the “a” coefficient), represents the fractal dimension:

$$a = D = \frac{n^2 \sum_{i=1,n} x_i y_i - \sum_{i=1,n} x_i \sum_{i=1,n} y_i}{n^2 \sum_{i=1,n} x_i^2 - \left(\sum_{i=1,n} x_i\right)^2}$$

(6)

where: $x_i=\log(1/s)$, and $y_i=\log(N(s))$.

The “box-counting” algorithm assumes to determine the fractal dimension according to the dependence of the object contour upon the used scale factor. It consists in successive image coverage with squares with equal sides (2, 4, 8, ...) and in counting every time the squares that cover the object contour.

The points of the coordinates ($\log(N(s))$, $\log(1/s)$), where $s$ is the common side of the coverage squares, and $N(s)$ the number of squares that contain any information, will be positioned approximately in a line and its slope will be the fractal dimension in the “box-counting” context. In a synthetic representation, the algorithm for determining the “box-counting” dimension for binary images is the following:

1. the original image (binary) is read;
2. the analysed region is selected;
3. the box-counting dimension is calculated by counting each time the number of cubes $N(s)$ that contain at least a point of the form. Logarithm is applied to the obtained values then, they are graphically represented by a curve with a slope that is the box-counting dimension.

For an example of how the algorithm is used, we’ll consider the image of a leaf (Fig. 4a) from which we’ll extract a binary version by neglecting all the pixels over a certain luminosity (Fig. 4b).

![Fig. 4. (a) The initial image; (b) The binary (black-white) image version](image-url)
Next, we’ll apply the “box-counting” algorithm, described above, for different scale values $s$ (only the squares that contain information are plotted, the ones that cover the leaf contour), by using an original software product, presented in detail in (Crisan, 2006).

![Figure 5. Object coverage with squares of different side values – “s”.](image)

We obtain the values table and “log-log” curve from the figure below:

![Figure 6. The log-log curve and the $s$, log$(1/s)$, log$(N(s))$ values.](image)

By using the least squares method, with the pairs of points (log$(N(s))$, log$(1/s)$), the regression line with the slope 1.55 is determined. Thus, the fractal dimension for the studied leaf is 1.55.

### 3. Case studies

In the presented case studies we will focus on the results of our original work in the fractal applications in botany and medicine.
3.1 Fractal analysis in botany
The actual application is dedicated to analysing the particularities of some species from the Gentianaceae family, with the purpose of establishing their affiliation to the Gentia genus. For establishing the independent position of the Gentianopsis genus, respectively the Gentianella ciliata for the Gentiana genus, we have fractally analyzed the species Gentiana lutea and Gentianella ciliata. The analysed material was acquired from the transversal section made through the root, stem and leaf of the mentioned taxons. The microscopical images of the extracted samples were analysed by using fractal techniques based on the “box-counting” algorithm in order to use other criteria than the morphological ones for establishing the position of the Gentiana, Gentianella and Gentianopsis genera in the Gentianaceae family.

3.1.1 Experimental processing
The Gentianaceae family was divided as time passed, in many genera (I. Prodan, 1939, Al. Beldie, 1979, Fr. Ehrendorffer, 1998, V. Ciocăran, 2000); from these the Menyanthes and Nymphoides genera came from the initial family and formed a new family named Menyanthaceae. At the moment, in the Romanian flora there are recognized 8 genera of the gentianacee’s family: Blackstonia, Centaurium, Comastoma, Gentiana, Gentianella, Gentianopsis, Lomantogonium, Sweertia (V. Ciocăran, 2000), obtained by the separation of the central genus Gentiana that has 13 species, into Gentianella with 5 species and Gentianopsis with one species. In the reference paper Flora Europaea (T. G. Tutin, 1972), the Gentianopsis genus is included at Gentianella under the name of Gentianella ciliata. Today, the Gentiana genus has 19 species spread all over our country. We are mentioning that the separation of the Gentianella and Gentianopsis genera was made on pure morphological criteria (Ciocarlan, 2000; Kinker, 1994).

For establishing the independent position of the Gentianopsis genus, respectively of the Gentianella ciliata for the Gentiana genus, we have fractally analyzed the species Gentiana lutea and Gentianella ciliata. Taking into account that the fractal technique is used as premiere work in this field of botany, for verifying its “sensibility”, we have compared the fractal dimensions of the 2 taxons from Gentianaceae with one from the Ranunculus genus from the Ranunculaceae family.

The analyzed material was acquired from the transversal section made through the root, stem and leaf of the mentioned taxons. The microscopical images of the extracted samples were analyzed by using fractal techniques based on the “box-counting” algorithm in order to use other criteria than the morphological ones for establishing the position of the Gentiana, Gentianella and Gentianopsis genera in the Gentianaceae family. Concretely, there were extracted window-images from the rind and the central cylinder of the root and stem and also from the mesophyll and nervure/rib of the leaf. The contours of the window-images were processed with the “box-counting” algorithm in order to establish the fractal dimensions for the analyzed sections.

The images have been analyzed by using an original software system described in detail in [5] which implements the extended box-counting algorithm for gray-level shapes. These images are binarized by using different thresholds, the contour is traced, and then the box-counting fractal dimension is represented into a spectrum. Finally, the representative fractal dimension (where the fractal character is preponderant) is selected for each tissue.
3.1.2 Results
For Gentiana lutea - root, the fractal dimension of the rind is $D_{11}=1.71$ and the central cylinder dimension is $D_{12}=1.80$:

Fig. 7. Fractal dimensions of the root, in the rind and central cylinder of Gentiana lutea.
For Gentiana lutea - stem, the fractal dimension of the rind is $D_{21}=1.63$ and the dimension of the central cylinder is $D_{22}=1.75$:

![Fractal dimensions for the stem in the rind and central cylinder for Gentiana lutea.](image)

Fig. 8. Fractal dimensions for the stem in the rind and central cylinder for Gentiana lutea.
For Gentiana lutea - leaf, the fractal dimension of the mesophyll is $D_1=1.55$ and the dimension of the vascular bundle is $D_2=1.69$:

Fig. 9. Fractal dimensions for the leaf, in the mesophyll and vascular bundle for Gentiana lutea.
For Gentianella cilliata (Gentianopsis) - root, fractal dimension of the rind is $D_{11}=1.65$ and the central cylinder dimension is $D_{12}=1.86$:

Fig. 10. Fractal dimensions for the root, in the rind and central cylinder for Gentianella cilliata (Gentianopsis).
For Gentianella cilliata (Gentianopsis) - stem, the fractal dimension of the rind is $D_{21}=1.62$ and the dimension of the central cylinder is $D_{22}=1.78$:

Fig. 11. Fractal dimensions for the stem in the rind and central cylinder for Gentianella cilliata (Gentianopsis).
For Gentianella cilliata (Gentianopsis) - leaf, the fractal dimension of the mesophyll is $D_{31}=1.58$ and the dimension of the vascular bundle is $D_{32}=1.77$.

Fig. 12. Fractal dimensions for the leaf, in the mesophyll and vascular bundle for Gentianella cilliata (Gentianopsis).
For Ranunculus repens - root, the fractal dimension of the rind is $D_{11}=1.55$ and the central cylinder dimension is $D_{12}=1.60$:

Fig. 13. Fractal dimensions for the root in the rind and central cylinder for Ranunculus repens.
For Ranunculus repens - stem, the fractal dimension of the rind is $D_{21}=1.52$ and the dimension of the vascular bundle is $D_{22}=1.57$:

Fig. 14. Fractal dimensions for the stem, in the rind and vascular bundle for Ranunculus repens.
For Ranunculus repens - leaf, the fractal dimension of the mesophyll is $D_{31}=1.53$ and the dimension of the vascular bundle is $D_{32}=1.58$:

Fig. 15. Fractal dimension for the leaf, in the mesophyll and vascular bundle for Ranunculus repens.
Table 1. Comparative table of the obtained data

In conclusion, the presented application, from biology is dedicated to analysing the particularities of some species from the Gentianaceae family, with the purpose of establishing their affiliation to the Gentina genus. To achieve this goal we have fractally analyzed microscopical images from the species of Gentiana lutea, Gentianella ciliata and Ranunculus repens.

The different positions of Gentianella ciliata (Gentianopsis) and Ranunculus repens can be observed. From our provisional observations it results that, from the fractal dimensions point of view (central cylinder - root, central cylinder - stem, vascular bundle - leaf), the separation of the Gentiana and Gentianopsis genera is justified.

The authors present for the first time in this field of botany a new method for differentiating some species of the Gentiana genus, by using the fractal analysis in order to establish the position of the Gentianella and Gentianopsis genera.

We consider that this paper opens new prospects in the botanical research domain and in biology in general.

3.2 Fractal analysis in medicine

Important results coming from the usage of fractal properties were obtained in medicine. Further on, some results in breast lesions classification will be presented.

Breast cancer is the most common women disease in modern world; statistics shows that a woman’s lifetime risk of developing breast cancer is 1/8. Mammography is the most efficient tool for the detection and diagnosis of breast lesions. In the last decades, medical exams became a regular act; thus, the amount of mammograms interpreted by a radiologist increased dramatically. As a result, a focused effort initiated two decades ago, is under way to develop a Computer-Aided Diagnosis of Mammograms (CADM).

One of the most important components in a CADM is to classify the lesion. The similarity between the breast tissue and synthetically generated fractals shown in fig. 16 (Sari-Sarraf et al., 1996) suggests that the fractal properties, such as the fractal dimension, may be used as a classifier.

Fractal dimension measures the complexity of an object; it grows as the shape is more irregular, as it can be seen in the table below. This observation will be very useful in order to characterize mammographic lesions.
Fig. 16. The similarity between the breast tissue (a) and synthetically generated fractals (b)

Table 2. The fractal dimension grows as the shape is more irregular

<table>
<thead>
<tr>
<th>Shape</th>
<th>Fractal dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$Df = 1$</td>
</tr>
<tr>
<td></td>
<td>$Df = 1.27$</td>
</tr>
<tr>
<td></td>
<td>$Df = 1.46$</td>
</tr>
</tbody>
</table>

3.2.1 BI-RADS classification

BI-RADS (Breast Imaging Reporting and Data System) is a very complex system proposed by The American College of Radiology (ACR) in order to classify mammographic lesions. The purpose of the BI-RADS system is to standardize mammography reporting in order to reduce confusion in breast imaging interpretations and facilitate outcome monitoring. BI-RADS system consists of five categories from 1 to 5; each of them characterizes a kind of mammographic lesion and implies a certain action as a treatment. Shortly, the five categories are:
• BI-RADS 1 – the category is referring to negative cases
• BI-RADS 2 – also describes a negative lesion, but in this case the interpreter may wish to describe a finding
• BI-RADS 3 – the third category refers to a probably benign finding, in this case a short interval follow-up is suggested
• BI-RADS 4 – characterizes the lesions that do not have the characteristic morphologies of breast cancer but have a definite probability of being malignant. In those cases, the radiologist may appeal to a biopsy.
• BI-RADS 5 – characterizes the lesions having a high probability of being cancer

3.2.2 Hypothesis and experiments
When categorizing a mammographic anomaly, the radiologist has to observe several properties of the lesion:
• the contour’s shape
• localization
• dimension
• density
• number and bolarity of anomalies
• presence or absence of associated microcalcifications.

One of the most important features is the contour’s shape: a regular contour is associated to a benign case, while an irregular shape characterizes a malign lesion. As table 2 shows, the fractal dimension grows with the irregularity of the shape; this could be an essential observation in order to classify the BI-RADS 4 lesions, with no need for further investigations or biopsy. The fractal dimension may provide a tool for classification: the lesions with a regular contour are more probably benign, while the lesions with an irregular contour are more probably malign.

A statistical experiment was developed on a lot of 30 cases. The hypothesis was tested on these cases of BI-RADS 4 classified lesions, 18 benign cases and 12 cancers provided by the Medical Imaging Department of Fundeni Clinical Institute, of Bucharest.

Each mammogram was analysed by following these steps:
Step 1 - the radiologist traces a FAR (Focussed Attention Region), by using a mobile cursor. The size area can be of 64X64, 128X128, 256X256 or 512X512. The selection must contain the anomaly and it is based on the radiologist’s experience. Budging the selection to the left or right, top or bottom will not influence the results of the analysis.

Fig. 17. A FAR traced by the radiologist.
Step 2 - the image is binarized by using a threshold between 1-255 gray level: all the pixels whose gray level is greater or equal to the threshold will be transformed into white pixels; the rest will become black. At this point, the forms inside the image are white on a black background.

![Image](image.png)

Fig. 18. The FAR is binarized; the white pixels are part of the form on a black background

Step 3 - the contour is automatically traced: once the image is binarized, the next step is to trace an outline of the white areas: all the white pixels which have at least one black neighbour will become part of the contour (every pixel has 8 neighbours: N, NE, E, SE, S, SV, V, NV). The rest of the pixels will be transformed into black pixels.

![Image](image.png)

Fig. 19. The contour is traced - an outline of the white areas.

Step 4 - the fractal dimension of the outline will be computed by using the box-counting algorithm. The result will be 1.36.
The box-counting algorithm will provide the 1.36- fractal dimension.

The results of 30 cases of BI-RADS 4 classified lesions are as follows: the benign lesions have lower fractal dimensions, between 1-1.50, while malign lesion have higher dimensions, between 1.35-2.

<table>
<thead>
<tr>
<th>Lesions</th>
<th>Fractal dimension</th>
<th>Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benign (18 cases)</td>
<td>&lt;1.4</td>
<td>16 (89%)</td>
</tr>
<tr>
<td></td>
<td>&gt;1.4</td>
<td>2 (11%)</td>
</tr>
<tr>
<td>Malign (12 cases)</td>
<td>&lt;1.4</td>
<td>1 (8%)</td>
</tr>
<tr>
<td></td>
<td>&gt;1.4</td>
<td>11 (92%)</td>
</tr>
</tbody>
</table>

Table 3. The fractal dimensions on 30 mammographic lesions

In fig. 21 it is presented the statistical result based on the fractal study. The fractal dimensions corresponding to 16 benign cases are situated below the 1.4 - threshold, while only two benign cases have higher dimensions; meanwhile, 11 malign cases have higher fractal dimensions, above 1.4 and only one case is below the 1.4-threshold.

Fig. 21. The fractal dimensions distribution on 30 mammographic lesions.
3.2.3 Conclusions
The presented application, from biomedicine, involves non-invasive techniques based on the processing of mammographic images. The method allows the diagnosing of mammographic tumours and it is based on two observations:

- the fractal dimension grows as the irregularity of the object grows;
- regular outline of a lesion is associated to a benign lesion, while irregular outline is associated to a malign lesion.

The hypothesis that cancers have higher fractal dimensions than benign lesions was tested on 30 cases and the results are encouraging.

4. References
Ehrendorfer Fr. (1998), „Lehrbuch der Botanik“, (Ed. 34,.) Gustav Fischer Verlag, Stuttgart
*** Breast Imaging Reporting Data System - http://www.birads.at
There are six sections in this book. The first section presents basic image processing techniques, such as image acquisition, storage, retrieval, transformation, filtering, and parallel computing. Then, some applications, such as road sign recognition, air quality monitoring, remote sensed image analysis, and diagnosis of industrial parts are considered. Subsequently, the application of image processing for the special eye examination and a newly three-dimensional digital camera are introduced. On the other hand, the section of medical imaging will show the applications of nuclear imaging, ultrasound imaging, and biology. The section of neural fuzzy presents the topics of image recognition, self-learning, image restoration, as well as evolutionary. The final section will show how to implement the hardware design based on the SoC or FPGA to accelerate image processing.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:
