COMPUTER AIDED DIAGNOSIS IN MAMMOGRAPHY BASED ON FRACTAL ANALYSIS

DAN POPESCU, MAXIMILIAN NICOLAE, DANIELA ALEXANDRA CRISAN, NICOLETA ANGELESCU
Department of Control and Computers
POLITEHNICA University of Bucharest
Splaiul Independentei nr. 313, Bucharest 6,
ROMANIA

Abstract: An important tool in medical diagnosis is digital images analysis and interpretation. The objective of the present paper is to demonstrate the utility of fractal analysis in theoretic decision for doubtful cases in malignity of mammary tumours. We present the fractal and textural features able to classify these tumours in malign or benign. Thus, the most important feature is the local connected fractal dimension calculated for edges extracted from binary images obtained with different binary thresholds. We describe a tumour classification algorithm based on the histogram of the local connected fractal dimension. In order to validate the described algorithm, experimental results from 30 images (mammography) are presented. At the work ending, some concluding remarks are presented also.

Key-words: fractal dimension, local connected fractal dimension, textural features, image segmentation, computer aided diagnosis, mammary cancer, tumour shape, image classification.

1 Introduction

Image processing and image based analysis techniques have played an important role in many important medical applications based on computer aided diagnosis. These applications involve the automatic image acquisition, primary image processing, segmentation and extraction of features from the image. Then the processed data are used for a variety of classification tasks, such as distinguishing normal tissue from abnormal tissue, growing processes, etc. Depending upon the particular diagnostic task, the extracted features for classification process have: morphological properties, colour properties, fractal properties or textural properties.

The textural properties computed are closely related to the application domain to be used. Thus, Sutton and Hall [1] discussed the classification of pulmonary diseases by texture features. They used three types of texture features to distinguish normal lungs from diseased lungs: a directional contrast measure, an isotropic contrast measure, and a Fourier domain energy spectrum. Harms, Gunzer and Aus [2] used a combination between texture features (micro-edges and size of textons) and colour features to diagnose malignancy in blood cells. In [3], the authors used textural features to estimate tissue scattering parameters in ultrasound images.

Recently, fractal geometry was used to investigate the relationship between the complexity of the epithelial/connective texture interface (as determined by fractal dimension) and the malignancy of the gastric tumour [4], [5].

In this paper, based on digital image processing and classification, we intend to investigate the possibility of breast cancer presence in doubtful cases. A mammography is classified in a BI-RADS category (Breast Imagining Reporting Data System) from 1 to 5. The 1-3 categories signify that the probability to be a malignant tumour is very small; the 5^{th} category means that the probability to be a malignant tumour is very high. In the 4^{th} category the malignancy risk is 5-50% and, in this case, a biopsy is necessary. The tumour aspect is opaque with blurred edges. Thus, the information

about malignity is concentrated in the tumour contour.

After image acquisition, a primary image processing (noise rejection, segmentation - in order to obtain binary image - and contour extraction) is necessary. The contour is reduced to a single pixel width in order to avoid the effects of the thickness of the contour [10]. The algorithms for image primary processing are the following (nonlinear type):

- noise rejection (median filter):

$b_2' = \text{Me}\{a_2, b_1, b_2, b_3, c_2\}$;

- segmentation :

$b_2' = 0$ if $b_2 < P$, and $b_2' = 1$ if $b_2 \geq P$, where P is a threshold chose by the operator;

- contour extraction:

$$b'_2 = b_2 \wedge \left(\bar{a}_2 \vee \bar{b}_1 \vee \bar{b}_3 \vee \bar{c}_2\right)$$

All these operators are local type and the elements are in a 3x3 neighbourhood of the central pixel b_2 (Fig.1). The result of each operator is noted by b_2'.

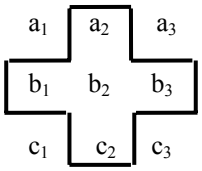

Fig.1.Local filtering cross type neighbourhood based.

2 Fractal dimension

In order to evaluate the malignity risk, one can analyse the tumour contour aspect. Thus, a method to relieve the irregularity of the contour is to calculate and combine different forms of fractal dimension.

There are many programs for counting the fractal dimension in different forms, but the most familiar algorithm is the box counting. Many specialized papers describe this algorithm.

One of the main characteristics of fractal objects is that their measure is depending on the scale used, fractal forms being not easily to measure in classic geometric context. Their physical properties (length, aria, volume) are depending on the representation resolution.

The *Hausdorff dimension* of a set F is defined as the most efficient covering of the set. Let consider $d,s \in R$ and a set of test function f(d) as:

$$N(s) = f(d) * s^d \qquad (1)$$

where N(s) is the smallest number of spheres of s-diameter needed to cover the given set F.

When $s \to \infty$, an unique value $d=D_H$ exists, called the Hausdorff dimension of F, so that:

$$d < D_H \Rightarrow N(s) \to \infty$$
$$d > D_H \Rightarrow N(s) \to 0$$

In equation (1) one can consider:

$$d = \lim_{s \to 0} \frac{\ln(N(s)) - \ln(f(d))}{\ln\left(1/s\right)}; \quad \lim_{s \to 0} \frac{\ln(f(d))}{\ln\left(1/s\right)} = 0 \qquad (2)$$

Then, the *fractal dimension* of set F is:

$$D_F = \lim_{s \to 0} \frac{\ln(N(s))}{\ln(1/s)} \qquad (3)$$

Although simple from theoretical viewpoint, the Hausdorff dimension is not easily to compute. Alternative methods are used, such as box-counting technique. The box-counting algorithm evaluates the fractal dimension, function of the evolution of the object size in relation with the scale factor used. This method is indicated in cases of homogenous structures, supplying global, average information of analyzed objects, because it ignores the heterogeneous nature of images. Such images, with different textures may have the same box-counting global dimension box. The existence, inside an image, of different regions with different fractal dimension requires alternative methods which will extend the fractal to multi-fractal notion.

Tumour growth is rather complex process, ultimately dependent on the presence of tumour cells that proliferate and spread in the host tissues. Most of the objects in biomedical images, like tumours, are multifractals. That means the fractal dimension has a variation in every point of the image according to the boundary of neighbourhood.

In order to describe the heterogeneous nature of a tumour image, we may compute, for every single point in the image, a local dimension (box-counting dimension, for instance), limited to a neighbourhood of the central pixel. Thus, instead of a single value meant to characterize the whole image, we have a set of values, one for each point in the analyzed object. The values will be represented into a histogram in order to give emphasis to the distribution of the local irregularities of the image. We may consider that the global dimension (of the

whole image) is the local dimension with the highest frequency.

In some situation neither the local approach of the fractal dimension is enough. A compact set of disconnected points (0-topological dimension) will have a higher fractal dimension. Thus, we associate to every single point in the image a local-connected fractal dimension, in order to describe the shape structure containing the point, considering only those points inside the neighbourhood connected with the central pixel.

Although the advantages of using the local and local connected dimensions are obvious, they present an important disadvantage: the distribution of the local and local connected fractal dimensions depends on the choice of the maximum window size.

3 Texture features used in computerized mammography

If a tumour image is analysed, we can observe different types of texture for malignant, benign or unclassified tissues. In order to discriminate different kinds of tumours, the theoretic decision method based on vector features extracted from texture is used. The main features which can be utilised for tumours classification are calculated from the co-ocurence matrix: correlation, energy, entropy, homogeneity and contrast.

A co-occurrence matrix is a two-dimensional array C in which both the row and the column running numbers represent a set of possible image values *V*. For grey-tone images *V* can be the set of possible grey tones and for colour images *V* can be the set of possible colours. The value of *C(i,j)* indicates how many times value i co-occurs with value j in some designated spatial relationship. For example, the spatial relationship may be that value i occurs immediately to the right of value j. To be more precise, we will look specifically at the case where the set *V* is a set of grey tones and the spatial relationship is given by a vector *d* that specifies the displacement between the pixel having value i and the pixel having value j.

Let *d* be a displacement vector *(dr, de)* where *dr* is a displacement in rows (downward) and *de* is a displacement in columns (to the right). Let *V* be a set of grey tones. The grey-tone co-occurrence matrix *Cd* for image *I* is defined by

$$C_d(i,j) = |\{((r,s),(t,v)): I(r,s) = i, I(t,v) = j\}|$$

where $(r,s), (t,v) \in N \times N$, $(t,v) = (r+dx, s+dy)$, $i,j \in V$, and $|*|$ is the cardinality of a set.

Figure 2 illustrates this concept with a 6 x 8 image *1* and two different co-occurrence matrices from *1:* $C_{(1,0)}$, and $C_{(-1,0)}$.

There are two important variations of the standard grey-tone co-occurrence matrix. The first is the *normalized* grey-tone co-occurrence matrix N_d defined by (4).

$$[I] = \begin{matrix} 0 & 0 & 1 & 1 & 2 & 2 & 3 & 3 \\ 3 & 3 & 0 & 0 & 1 & 1 & 2 & 2 \\ 2 & 2 & 3 & 3 & 0 & 0 & 1 & 1 \\ 1 & 1 & 2 & 2 & 3 & 3 & 0 & 0 \\ 0 & 0 & 1 & 1 & 2 & 2 & 3 & 3 \\ 3 & 3 & 0 & 0 & 1 & 1 & 2 & 2 \end{matrix}$$

$$[C_{(1,0)}] = \begin{matrix} 0 & 0 & 0 & 10 \\ 10 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 \\ 0 & 0 & 10 & 0 \end{matrix} \qquad [C_{(-1,0)}] = \begin{matrix} 0 & 10 & 0 & 0 \\ 0 & 0 & 10 & 0 \\ 0 & 0 & 0 & 10 \\ 10 & 0 & 0 & 0 \end{matrix}$$

Fig. 2. Two different co-occurrence matrices for a grey-tone image [*I*].

$$N_d(i,j) = \frac{C_d(i,j)}{\sum i \sum j Cd(i,j)} \tag{4}$$

This relation normalizes the co-occurrence values to lie between zero and one and allows them to be thought of as probabilities in a large matrix. The second is the *symmetric* grey-tone co-occurrence matrix $S_d(i,j)$ defined by

$$S_d(i,j) = C_d(i,j) + C_{-d}(i,j) \tag{5}$$

which groups pairs of symmetric adjacencies.

Co-occurrence matrices capture properties of a texture, but they are not directly useful for further analysis, such as comparing two textures. Instead, numeric features are computed from the co-occurrence matrix that can be used to represent the texture more compactly. The following are standard features derivable from a normalized co-occurrence matrix:

$$Energy = \sum_i \sum_j N_d^2(i,j) \tag{6}$$

$$Entropy = -\sum_i \sum_j N_d(i,j) \log N_d(i,j) \tag{7}$$

$$Contrast = \sum_i \sum_j (i-j)^2 N_d(i,j) \quad (8)$$

$$Homogeneity = \sum_i \sum_j \frac{N_d(i,j)}{1+|i-j|} \quad (9)$$

$$Correlation = \frac{\sum_i \sum_j (i-\mu_i)(j-\mu_i) N_d(i,j)}{\sigma_i \sigma j} \quad (10)$$

where μ_i, μ_j are the means and σ_i, σ_j are the standard deviations of the row and column sums $N_d(i)$ and $N_d(j)$, which are defined by relations (11),(12).

$$N_d(i) = \sum_j N_d(i,j) \quad (11)$$

$$N_d(j) = \sum_i N_d(i,j) \quad (12)$$

For the textural images the colour and the texture are more important of perceptual viewpoint because there are not group of objects.

The regions of textural images tend to spear in whole image, in time that the non-textural images are usual partition in group regions.

4 Algorithm for mammary cancer diagnosis. Experimental results

The outline of each image was analyzed by estimating the global fractal dimension, the local fractal dimension and local connected fractal dimension. For this purpose we used an original software package described in detail elsewhere [6]. In brief, the fractal dimension of each outline was measured by the box-counting algorithm and the local fractal dimension and local connected fractal dimension were estimate according to the algorithms published in [8] and [6].

The images were analyzed by estimating their local mass scaling properties. The computer program measured the total number of pixels locally connected in a window of increasing size, centred at a point [4]. Locally connected relates to all pixels within the largest box used which belong to the cluster connected to the pixel where the box is centred.

The algorithm for this procedure is:

1. Consider the current point P;
2. Mark all the points connected with P within a growing s-size window centred at P (s is under a fixed s_{max} value which may not be modified during the analysis of the whole image: 32, for example). Notice that every point has eight neighbour pixels (N, NW, V, SW, S, SE, E, NE).
3. Count every time how many points N(s) of the analyzed object are within the window;
4. Using the least square method, compute the slope of the log-log curve composed by the (ln(N(s),ln(s)) points.

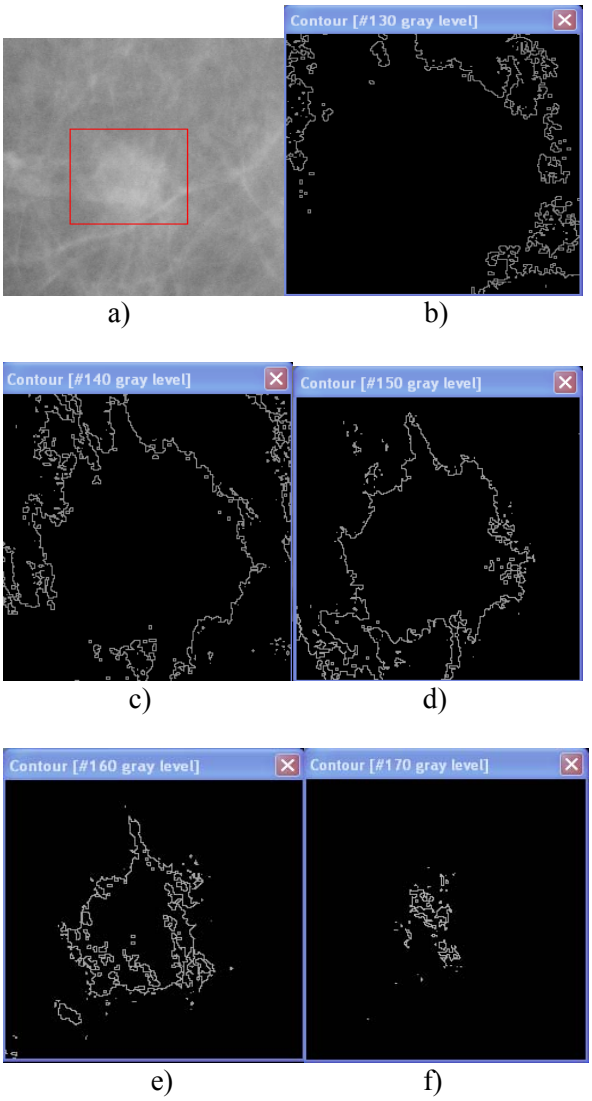

Fig.3. Mammary tumour BI-RADS 4. a) Initial image, 256 grey levels; b)130 threshold segmentation; c)140 threshold segmentation; d)150 threshold segmentation; e)160 threshold segmentation; f)170 threshold segmentation.

Most of the information about the malignity of a tumour is contained in the contour of the tumour shape. Doubtful tumours are characterized by blurred contours which are changing by different threshold used to separate the tumour from background segmentation. The image was 1024x1024 pixels, 256 grey levels, bmp format, with

a window 120 – 175 grey level for tumour representation. Different contours, for different thresholds are represented in the figures 3b, 3c, 3d, 3e, and 3f. Fig 3a represents the initial image corresponding to the mammography and the area of interest selected by the specialist. We can observe that the contours are different for different binary thresholds: 130, 140, 150, 160, and 170. The box counting fractal dimension spectrum is nearly constant in the domain 130-160 (Fig.4). Also the maximum frequency for local connected fractal dimension Df: 1.42, 1.38, 1.33, 1.36 (Fig.4, Fig.5, Fig.6, Fig.7), fluctuates very little.

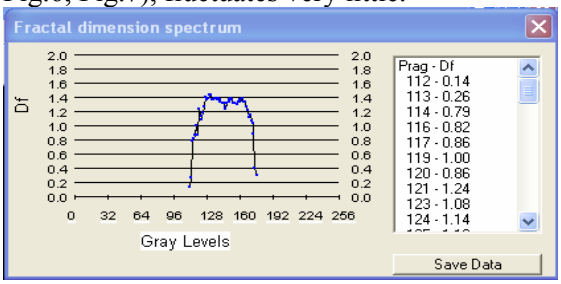

Fig.4.Box counting fractal dimension

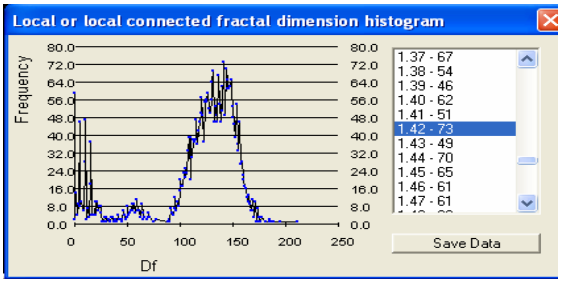

Fig.5. Local connected fractal dimension for Fig.3b, Df=1.42.

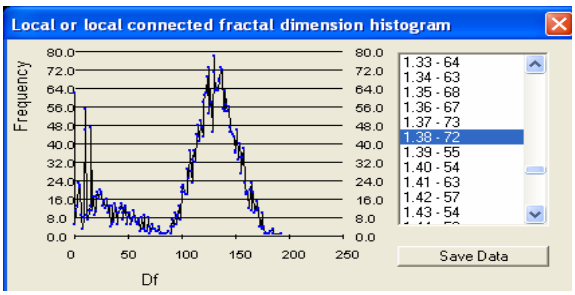

Fig.6. Local connected fractal dimension for Fig.3c, Df=1.38.

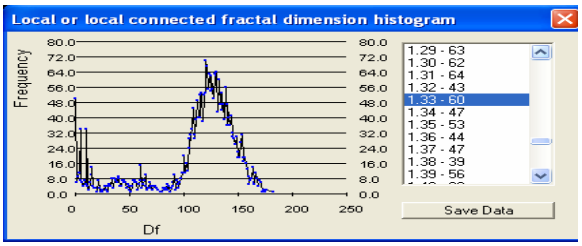

Fig.5. Local connected fractal dimension for Fig.2d, Df=1.33.

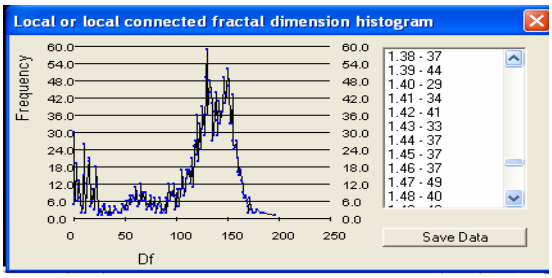

Fig.8. Local connected fractal dimension for Fig.3e, Df=1.36.

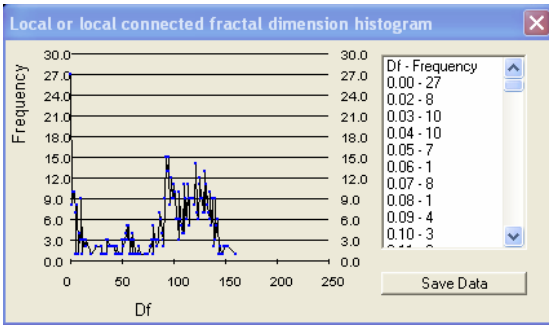

Fig.9. Local connected fractal dimension for Fig.3f, Df=1.12.

A set of 30 images (mammography BI-RADS 4), from Fundeni Clinical Institute, Bucharest, was processed and, in each case, the medium local connected fractal dimension was calculated, for different segmentation thresholds. The experimental values are presented in Table1 and in Fig.10 and a diagnosis threshold equal to 1.4 in medium fractal dimension were established. A medium fractal dimension less than 1.4 represents a benign case and a medium fractal dimension exceed 1.4 represents a malignant case. After tumour evolution investigation 18 cases was classified like benign and other 12 cases was classified like malignant. The percentage of correct diagnosis in the malignant case was approximately 92%.

Table 1. Experimental values

	Benign (18 cases)		Malign (12 cases)	
Medium fractal dimension	<1.4	>1.4	<1.4	>1.4
Cases	16 (89%)	2 (11%)	1 (8%)	11 (92%)

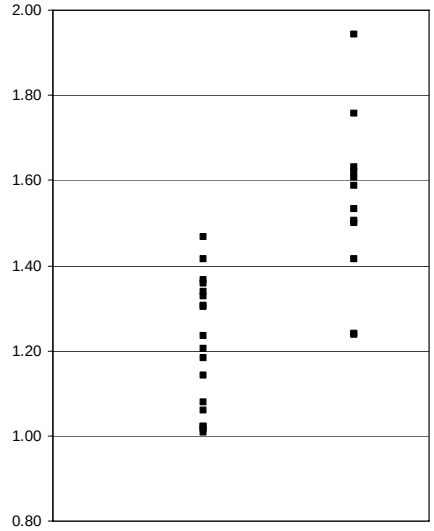

Fig.10. Mean of the fractal dimension: left- benign cases, right- malignant cases.

References

[1] Sutton, R.; E. L. Hall, Texture Measures for Automatic Classification of Pulmonary Disease, *IEEE Transactions on Computers*, vol.C-21, 1992, pp. 667-676;

[2] Harms, H., U. Gunzer, and H. M. Aus (1996), Combined Local Color and Texture Analysis of Stained Cells, *Computer Vision, Graphics, and Image Processing*, vol.33, 1996, pp.364-376;

[3] Insana, M. F., R. F. Wagner, B. S. Garra, D. G. Brown, and T. H. Shawker, Analysis of Ultrasound Image Texture via Generalized Rician Statistics, *Optical Engineering*, vol.25, 1996, pp. 743-748;

[4] Dobrescu, R., M. Dobrescu and F. Talos, Multifractal medical image analysis using fractal dimension, in: R. Dobrescu, C. Vasilescu (Eds) *Interdisciplinary Applications of Fractal and Chaos Theory*, Editura Academiei Romane, Bucuresti, 2004;

[5] Vasilescu, C., A. Herlea, B. Ivanov, R. Dobrescu, F. Talos (2004), A survey on differences between intestinal and diffuse type of gastric carcinoma, in : R. Dobrescu, C. Vasilescu (Eds) *Interdisciplinary Applications of Fractal and Chaos Theory*, Editura Academiei Romane, Bucuresti, 2004;

[6] Crisan, A.D., *Image processing using fractal techniques*, Ph.D Thesis, Politehnica University Bucharest, 2005;

[7] F.Talos, Using local fractal dimension in multifractals analysis, *Proceedings of the "Interdisciplinary approaches in fractal analysis*", IAFA 2003, Bucharest, Romania, 2003;

[8] G. Landini and J.W. Rippin, How important is tumor shape? Quantification of the epithelial-connective tissue interface in oral lesions using local connected fractal dimension analysis, *J. Pathol.*, 179, 1996, pp210-217;

[9] Barnsley, M., *Fractals Everywhere*, Academic Press, 1988;

[10] Popescu, D., Dobrescu, R, Mocanu, S., *Dedicated Primary Image Processors For Mobile Robots*, WSEAS Trans. on Systems, Issue 8, Vol.5, August 2006, p. 1932-1939;