IMAGE ANALYSIS BASED ON THE STUDY OF THE ATTRACTOR OF A TIME SERIES

Andreea UDREA[1], Mircea OLTEANU[2]

Lucrarea are ca scop optimizări ale metodei seriilor de timp pentru analiza de imagini, cu aplicații pentru clasificarea texturilor. Dimensiunea de corelație a atractorului seriei de timp este criteriul de caracterizare. În această lucrare sunt prezentate şi argumentate metodele şi implementările lor. Concluziile studiului se bazează pe o analiză statistică a 50 de imagini TC (tomografii computerizate).

The paper proposes some improvements of nonlinear time series methods for image analysis; used for texture classification. The correlation dimension of the time series attractor is the criterion for characterization. In this paper we present and argument the methods and their implementations. The conclusions of the study are based on the statistical analysis of 50 CT images (computer tomographies).

Keywords: spatial series, attractor's correlation dimension, texture classification

1. Introduction

Nonlinear time series analysis and fractal analysis, branches of chaos theory, provide useful methods for the characterization of single and multi variable signals (time series, images).

Typically, nonlinear time series analysis deals with signals that are sets of values of a single variable function, usually measured as function of time (*dynamic features*). Nonlinear methods were developed in the last 20 years, motivated by the concept of deterministic chaos which was proved to exist within many real systems in chemistry, physics, biology, medicine, electronics. The studied time series are: recordings of the electrical activity – electrocardiograms [1], electroencephalograms [2] and physiological parameters – blood pressure , breathing [3]; voice signal [4], laser output signal [5], meteorological parameters values [6], etc .

On the other hand, fractal analysis methods are used for the description and classification of *geometric features* of irregular forms and patterns. Its most known tool is the fractal dimension used to provide information on the irregularity of the contour of an object or selfsimilarities of a texture. It was largely applied

[1] Assist., Faculty of Automatic Control and Computers, University POLITEHNICA Bucharest, Romania

[2] Prof., Dept. of Mathematics I, University POLITEHNICA Bucharest, Romania, e-mail: mirolteanu@yahoo.co.uk

for the study of biological systems and subsystems at microscopic and macroscopic scale, image enhancement and compression, fracture analysis, texture classification ([7], [8], [9]).

In this paper we describe and test a methodology derived from the time series analysis, which, by means of studying the attractor of a spatial series associated to an image or image region, provides information that can be used for image classification and characterization.

2. Nonlinear time series analysis – theoretical background

In this section we present a few basic results and methods on nonlinear time series analysis.

Let *(M, d)* be a complete metric space (*M* - is the set of states and *S* - the time set) and let *(S, +)* be an abelian semigroup.

A dynamical system on *M* is an arbitrary map $T: S \times M \to M$ such that:

$$T(0,x) = x, \ \forall x \in M$$

$$T(t,T(s,x)) = T(t+s,x), \ \ \forall t,s \in S \ ,\forall x \in M$$

A dynamical system $T: N \times M \to M$ is said to be a discrete dynamical system if there is a map $f: M \to M$ such that:

$$T(n,x) = (f \circ f \circ \ldots \circ f)(x) = f^n(x),$$
$$\forall n \in N, \forall x \in M.$$

A dynamical system *T* is called chaotic if:

- *T* has a sensible dependence on the initial condition on *M*
- *T* is topologic transitive on *M*
- the periodic orbits generated by *T* are dense in *M*.

A nonempty set of states $K \subset M$ is called an attractor or attracting set for the system *T* if the following properties hold:

1. – *K* is closed
2. – *K* is *T*- invariant,
3. – there is a neighborhood *U* of *K* such that:

$$\lim_{t \to \infty} d(T(t,x),K) = 0, \ \ \forall x \in U \cdot$$

The largest open set *U* satisfying the third condition is called the basin of attraction of *K*. An attractor is said to be global if the basin of attraction is the entire *M*.

The attractor of a chaotic system is called a strange attractor if it has a non-integer fractal dimension.

A real valued map

$$F: M \to R$$

is interpreted as a physical measure on the state space. If $\forall t, s \in S$ are fixed (*s* is called delay) and $x \in M$ is a fixed state, then a sequence of measurements:

$$F(T(t,x)), F(T(t+s,x)), F(T(t+2s,x)), \ldots, F(T(t+(d-1)s,x)), \ldots$$

is called a time series starting from *(t, x)* associated to the system *T*.

If *T* is a discrete dynamical system defined by the map *f*, then the associated time series starting from *(0, x)* is:

$$F(x), F(f(x)), F(f^2(x)), \ldots, F(f^n(x)) \ldots$$

By investigating the time series associated to a dynamical system, one can observe the behavior of that system. More precisely, by using Taken's Embedding Theorem ([10]) one can reconstruct, in an appropriate embedding space, the attractor from the time series generated by the system. The theorem is stated for an infinite time series, but for practical applications it can be implemented for a long enough signal sample.

Bellow we present a version of Taken's theorem due to T. Sauer, J.Yorke, M.Casdaglia ([11]).

Let $T: R \times M \to M$ be a smooth dynamical system of class C^2 on *M* and let $F: M \to R$ be a map (physical measurement) of class C^2. Let $t \in R$ be a fixed moment and let $\tau > 0$ be a time delay. If *K* is a compact invariant set of *T* and if *b* is the box-counting dimension of *K*, then the map:

$$H: K \to R^{2b+1}$$

defined by:

$$H(x) = (F(T(t,x)), F(T(t-\tau,x)), \ldots, F(T(t-2b\tau,x)))$$

is generically injective (a property is called generic if it holds on a set which contains a countable intersection of open dense sets), hence it is an embedding of the attractor in the space R^{2b+1}.

Presuming that the fractal dimension of the attractor is known, the attractor can be reconstructed from a univariate time series in a higher dimensional space ([12]- [14]).

The correlation integral is defined by the following expression:

$$C(\varepsilon) = \lim_{N \to \infty} \frac{1}{N^2} \sum_{i,j=1}^{N} H(\varepsilon - |y_i - y_j|) \qquad (1)$$

where:

- *H(x)* - is the Heaviside function, $H(x) = \begin{cases} 1, x > 0 \\ 0, otherwise \end{cases}$;
- ε - accepted distance between points;
- y_i - is a point in the embedded phase space constructed from a single time series according to Taken's theorem:

$$y_i = (x_i, x_{i+\tau}, x_{i+2\tau}, \ldots, x_{i+(d_E-1)\tau});$$

- d_E - the dimension of the embedding space;
- τ - the time delay;
- $i = N-\tau\,(d_E+1)$ – number of embedding vectors;
- N - initial time series length.

So, $C(\varepsilon)$ gives the proportion of the number of pairs of points in the embedding space with the Euclidian distance less than a specified small ε.

The dependence between the correlation integral and the correlation dimension is given by:

$$C(\varepsilon) = \varepsilon^{d_C(\varepsilon)} \tag{2}$$

The correlation dimension of the attractor is calculated using the formula:

$$d_C = \lim_{\varepsilon \to 0} \frac{\ln C(\varepsilon)}{\ln \varepsilon} \tag{3}$$

The d_C for a closed curve is 1, for a two-dimensional surface is 2, while for a strange attractor is a non-integer number.

The autocorrelation function of a signal x of length N is given by the following expression:

$$RN(\tau) = \frac{\sum_{n=1}^{N-\tau}\left(x_{n+\tau} - \bar{x}\right)\left(x_n - \bar{x}\right)}{\sum_{n=1}^{N-\tau}\left(x_n - \bar{x}\right)^2}; \quad \bar{x} = \frac{1}{N}\sum_{n=1}^{N} x_n \tag{4}$$

3. Image analysis

The time series associated to an image

In order to perform nonlinear analysis on an image, a number of steps must be made. First, the region of interest must be isolated from the image. The selected region is, in fact, a matrix -A- containing values of each pixels' shade (the value can vary between 0 and 255 corresponding to different shades of grey; 0 stands for black and 255 for white):

$$A = (a_{ij})_{\substack{i=1,n \\ j=1,m}}; \quad \forall a_{ij} \in \{0,\ 1,\ \ldots,\ 255\}$$

The time (spatial) series is generated in the following manner [17]-[19]:

1. The matrix obtained from the original image is cut in horizontal strips of 1, 4, 8, … pixels. The strip's width is a parameter chosen by the analyst by taking into account of the image resolution (for low resolutions the strip's width must be one pixel).
2. All strips are put together one after another and generate one single strip.

3. The time (spatial) series - *x(t)* - is generated by computing either the mean value or the maximal (dominant) value of each column of pixels within the strip.

4. The series is normalized by dividing each value at the domain variation, namely the grey degrees (256 shades).

Steps 1-2 stand for applying the *vec* operator to the image. *Vec* is a linear operator. It creates a column vector from a matrix *A* by stacking the column vectors of *A* below one another:

$$vec: A_{n,m}(R) \to R^{n \cdot m}$$

$$vec(A) = col(a_{i1},\ a_{i2}, \ldots,\ \mathrm{a}_{\mathrm{im}}), i = \overline{1,n};$$

The correlation dimension is invariant to smooth transforms so by applying *vec* to the image matrix its value is unchanged.

As result of this procedure, the time (spatial) series associated to the section of the analyzed tissue is obtained (Fig. 1.).

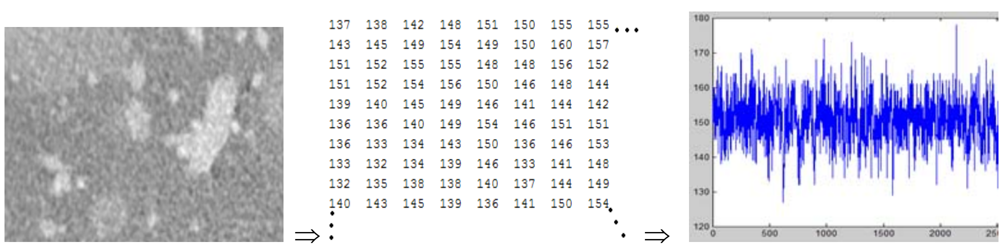

Fig. 1. Liver CT image region, the image matrix and the associated time series (position in the strip vs. grey level)

Having the associated series, the next step of the procedure implies the computation of the correlation dimension of the attractor. This value will be used as a discrimination criterion.

However, in practical applications, in order to determine the d_C, we can not directly use the above formulae due to the following aspects: the measurements are finite sequences of values (they are not time series in the mathematical sense), the measurements are corrupted by noise, the fractal dimension of the attractor is unknown and for different τ - delay values one obtains different results.

That is why specific methods for determining the delay, the suitable embedding space and the correlation dimension were refined over the years. Here are given the outlines for these methods together with our improvements. They were implemented as a new MatLab nonlinear analysis toolbox dedicated to image analysis.

The procedure for the time delay

The delay or lag value, τ, used to create the delayed embedding must be chosen carefully. A small value of the delay generates correlated vector elements and the geometry in the reconstructed phase space is stretched out along the diagonal (Fig. 2.a). A large value chosen for the delay generates uncorrelated vectors and a random distribution of points in the embedding space (Fig. 2.b).

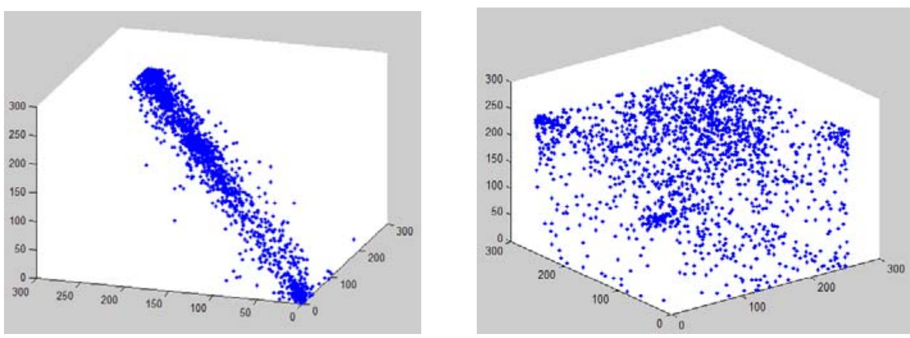

Fig. 2. Attractor in embedding dimension 3; the delay for the embedding is a) too small τ =2 – correlated vectors; b) too large τ =20 – random distributed vectors

According to [16] an appropriate delay, τ, can be chosen with good results (Fig. 3.b) if the autocorrelation function of the reconstructed series decays with 1/e of its initial value (Fig. 3.a):

$$RN(\tau) < RN(1)(1-1/e)\,. \qquad (5)$$

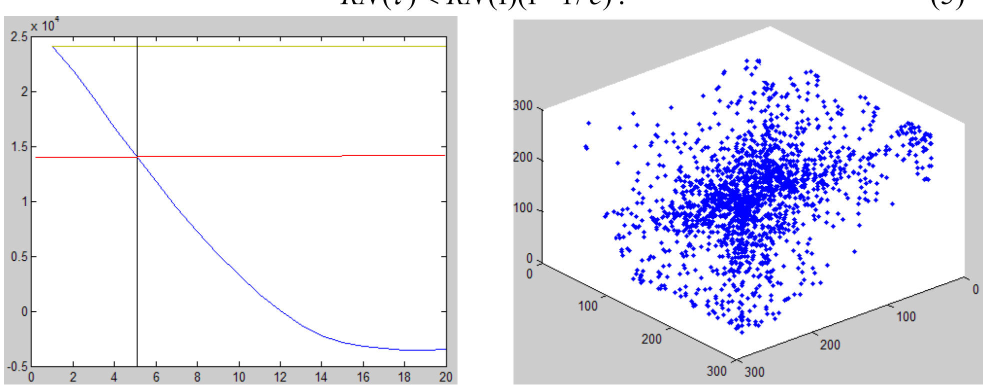

Fig. 3. a) delay value vs. autocorrelation function; b) reconstructed attractor in embedding dimension 3 with delay τ =5

In order to determine this value, the MatLab routine searches over a specified range, given by the user $[1, \tau_{\max}]$ for the first value that satisfies relation (5). Generally, the lag value (delay) is found between 3 and 10.

The procedure for the embedding dimension

The minimum allowed embedding dimension is the dimension where the number of so called false nearest neighbors (concept introduced by Kennel, Brown & Abarbanel (1992)) drops under a certain percent. A false neighbor is a point that under a certain higher dimensional embedding is projected near a point that in the previous embedding was not in its vicinity.

In order to implement this procedure, each point of the delayed series is tested by taking its closest neighbor in d_E dimensions, and computing the ratio of the distances between these two points in $d_E +1$ dimensions and in d_E dimensions. If this ratio is larger than a certain threshold ε_{d_E} see relation (6), the neighbor was false (this threshold is taken large enough to take in consideration points that exponential divergence due to deterministic chaos):

$$\frac{\left\| y_{i,d_{E+1}} - y_{j,d_{E+1}} \right\|}{\left\| y_{i,d_E} - y_{j,d_E} \right\|} > \varepsilon_{d_E} \tag{6}$$

where ||.|| is the Euclidian distance.

The MatLab routine calculates the percentage of false neighbors over a range of embedding dimensions (d_E between 2 and 15) and until it reaches a value less than a specified limit; otherwise it considers the minimal obtained value.

Once a proper delay and a minimum allowed embedding dimension are determined, the correlation dimension is calculated over a range of different ε -values and embedding dimensions higher than the first assuring a decreased number of false neighbors. The trusted interval for ε is searched generally between 0.001 and 1. The double logarithmic linear region between ε and the correlation integral is used for computing the d_C value.

Also, the d_C differs from one embedding dimension to another due to noise in the data, but there is a particular region, usually called the *scaling region* where d_C stabilizes ([16]). There is the interval where a mean value for the correlation dimension of an attractor is calculated.

4. Results and statistical analysis

In order to demonstrate if the above methodology can be applied with promising results for pattern classification, we analyze a set of 50 CT images containing normal and modified liver tissue.

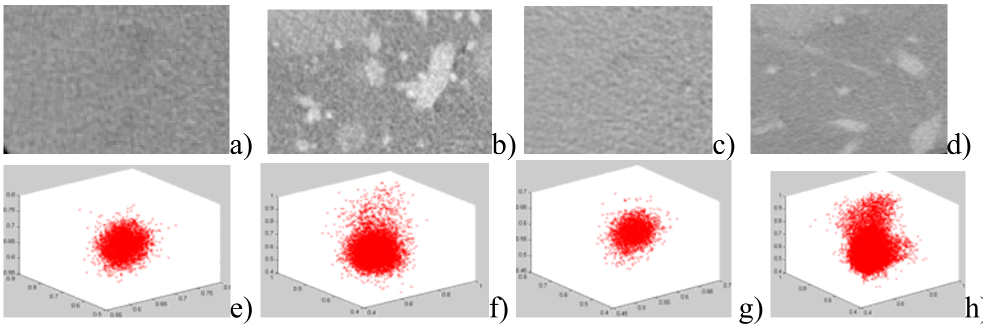

Fig. 4. Liver regions selected from CT image with contrast substance a),c) – normal; b), d) malign and their reconstructed attractors in e) - h)

In Fig. 4 two pairs of samples (normal and modified tissue) are presented. It can be observed that the attractors have distinct structures: connected - for normal tissue and disconnected (sometimes with several different attracting regions) - for the modified tissue.

In order to determine the ε-interval, we plotted the correlation dimension vs. the logarithm from ε (Fig 5.) and determined the intervals: for the normal tissue –$I_1=[0.001, 0.007]$ and for the modified tissue – $I_2=[0.003,0.01]$. In order to process the images automatically, the ε- interval is chosen as the intersection between I_1 and I_2: $I=[0.003,0.007]$. Using these values the correlation integral is determined.

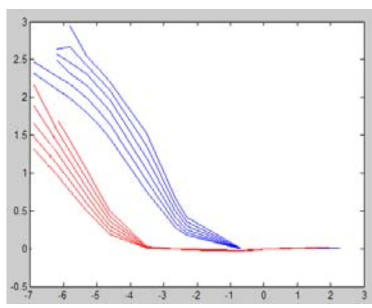

Fig 5. In red (normal tissue) and in blue (modified tissue) log(ε) vs. correlation dimension plot

The correlation dimension was calculated over a set of embedding dimensions and the determined scaling region was found in the interval [11-14]. For our study, in order to automate the process d_E was chosen 13.

The statistical analysis was done on a set of 50 CT liver images, divided into two samples (normal and modified).

For the statistical analysis, descriptive and comparison procedures were performed. For the correlation dimension of each sample, the average, standard deviation, standard skewness and standard kurtosis are computed.

The d_C confidence interval for normal tissue is [1.2920, 1.5217] and for modified tissue [1.8248, 2.8181].

Table 1.

Descriptive statistical methods results

Descriptive methods	Normal tissue d_C	Modified tissue d_C
Average	1.39683	2.4074
Stnd. Deviation	0.0828342	0.294393
Stnd. Skewness	0.348448	1.10369
Stnd. Kurtosis	-0.859346	0.582716

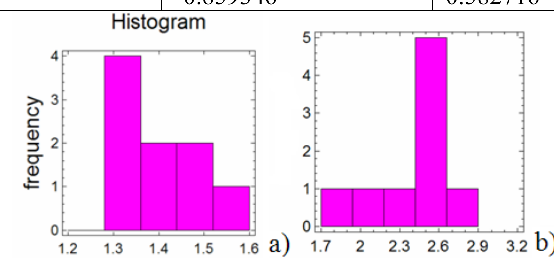

Fig. 6. Frequency histogram of the two samples a) normal b) modified

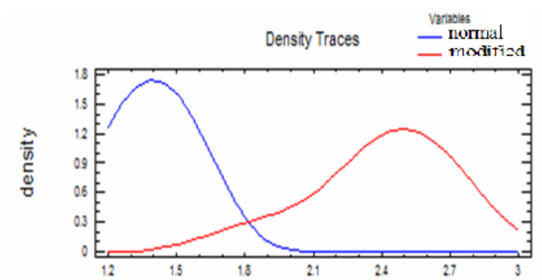

Fig.7. Comparison of the d_C density traces

By means of statistical analysis - Kolmogorov-Smirnoff test - we have compared the two distributions. Since the P-value is 0.00024682 (less than 0.05), there is a statistically significant difference between the two distributions at the 95.0% confidence level.

5. Conclusions

This paper presents some improvements towards texture classification based on nonlinear time series methods; it can be applied to spatial series obtained from the studied medical images. The differences between textures are captured by the correlation dimensions of the attractors of the spatial series. The applicability of the proposed method was demonstrated on a set of CT images containing normal and modified liver tissue. The method captures the differences between the two samples.

The statistical analysis reveals significant differences between the correlation dimension of the normal tissue and the correlation dimension of the modified tissue.

Future work aims at enlarging the liver CT images data base; and also apply the procedure to other kind of tissues.

R E F E R E N C E S

[1] *M. Perc*, Nonlinear time series analysis of human electrocardiogram, European Journal of Physics, 26, 2005, pp. 757-768

[2] *A. Mekler*, Calculation of EEG correlation dimension: Large massifs of experimental data, *Computer Methods and Programs in Biomedicine*, 92, 2008, pp. 154-160

[3] *H. Suyama, N. Burioka, T. Sako, M. Miyata, E. Shimizu*, Reduction of correlation dimension in human respiration by inhaling a mixture gas of 5% carbon dioxide and 955 oxygen, *Biomedicine & Pharmacology*, 57, 2003, pp. 116-121

[4] *V. Pitsikalis, P. Maragos*, **S**peech analysis and feature extraction using chaotic models, Proc. Int'l Conf. Acoustics Speech and Signal Processing (ICASSP-2002), Orlando, USA, May 2002. pp.533-536

[5] *M. Ciobanu, D. Savastru, S. Miclos, M. Tautan, .M. Rusu, S. Dontu*, Chaotic behavior of ideal for-level laser with periodic pump modulation, UPB Sci. Bull. Series A, 72, 2010,

[6] *M. A. Ghorbani, O. Kisi, M. Aalinezhad*, A probe into chaotic behavior nature of daily streamflow time series by correlation dimension and largest Lyapunov methods, *Applied Mathematics Modelling*, 2010

[7] J. *B. Bassingthwaighte, L.S. Liebovitch, B.J. West*, Fractal Physiology, Oxford University Press, 1994

[8] *G.Landini, G. A. Losa*, Complexity in Tumor Growth Patterns in Fractals in Biology and Medicine, Birkhauser Verlag, 1998, pg. 268-284

[9] *R. Dobrescu, C. Vasilescu*, Interdisciplinary Applications of Fractal and Chaos Theory, Academia Română, Bucuresti, 2004

[10] *F. Takens:* "Detecting strange attractors in turbulence", Dynamical systems and turbulence, Lectures Notes in Mathematics, 898, 1981, pp.366-381.

[11] *T. Sauer, J. A. Yorke, M. Casdagia*, Embedology, Journal of Statistical Physics, 65, 1991, pp. 579-616

[12] *P. Grassberger, I. Procaccia*, Characterization of strange attractors, Phys. Rev. Lett., 50, pp. 346-349

[13] *P. Grassberger, I. Procaccia*, Measuring the strangeness of strange attractors, Physica, 9D, 189-208

[14] *M. Gadea, C.P. Niculescu*, Chaotic Dynamical Systems, Universitaria Press, 2001.

[15] *R. Hegger, H. Kantz, T. Schreiber*, Practical implementation of nonlinear time series methods: The TISEAN package, *Chaos*, 9, 1999, pp. 413-435

[16] *H. Kantz, T. Schreiber*, Nonlinear Time Series Analysis, University Press, Cambridge, 2004

[17] *T. Mattfeldt, H. Kestler*, Classification of pancreatic tissue texture on the basis of correlation dimensions and Lyapunov exponents using two types of artificial neural networks, Interdisciplinary applications of fractal and chaos theory, Ed. Acad. Romane, 2004, pp. 225-240

[18] *M. Olteanu*, A nonlinear analysis study of time series associated to CT liver images, MCPL 2007, 4-th IFAC Conference , sept 2007 Sibiu, vol II, pp. 353-359

[19] *A. Udrea, L. Costache, M. Olteanu, D. Popescu*, On the relevance of nonlinear analysis of time series associated to CT kidney images, Proceedings of 2010 IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR 2010), Cluj-Napoca, 28-30 May, vol. 2, 2010 ,pp. 304-310