Fractal Random Walk and Classification of ECG Signal

Daoming Zhang, Guojun Tan, Jifei Hao
College of Information and Electrical Engineering
China University of Mining and Technology
Xuzhou, Jiangsu Province, China
zdmcumt@163.com

Abstract

This paper presents a new nonlinear method to analyze ventricular arrhythmia(VA) and sinus rhythms(SR). The problem is introduced from the discussion of Fractal Random Walk characteristics of ECG signal. Further, the fractal analysis is used to distinguish ventricular flutter(VFL), ventricular fibrillation(VF), ventricular tachycardia(VT)) and sinus rhythms(SR) from the raw electrocardiogram(ECG) data. The method has a three step processing. First, calculating the slope of permutation entropy(PE) to detect the onset of ventricular arrhythmia; Second, using regularization dimension(RD) to classify SR, VFL and VT/VF; Finally, according to multifractal spectrum(MS) area to distinguish VT and VF. Four databases are used to detect the method, and the accuracy of every step is 93.33%, 100% and 98%. As a whole, the accuracy of detecting onset of ventricular arrhythmia and confirming which ventricular arrhythmia is, is VFL 93.33%, VT and VF 91.47%.

1. Introduction

As is well known, ventricular arrhythmia is a very dangerous cardiac disease, even can make patient sudden death. So how to detect ventricular arrhythmia become very important. Ventricular arrhythmia(VA) has three kind: ventricular flutter(VFL), ventricular fibrillation(VF) and ventricular tachycardia(VT), many scholars develop their methods to detect various VT and/or VF [1]–[4]. All these methods have their advantages, but they mostly concentrate on time-frequency characteristic to analyze ventricular arrhythmia, and the few [5], [6] research the nonlinear behavior of ventricular arrhythmia. Also they scarcely ever make an evaluation about VFL, which is a kind of dangerous cardiac disease; it often appears before VF then rapidly changed into VF. So, detecting VFL precisely has also important clinical significance.

Hurst developed the re-scaled range analysis in investigating long time records of some natural processes [7]. As a good nonlinear research tool, scholars used it to explore the random walk characteristics of stock market and got many useful result, but no one put it on ECG signal analysis [8], [9]. In this paper, we introduce it to sinus rhythms and ventricular arrhythmia analysis, and find that Hurst exponent(Hurst dimension) has evident difference. To make the difference clear, we study further the nonlinear characteristics of these two kind signals.

Since Goldberger [10] and Ivanov [11] put forward fractal mechanisms in the electrophysiology of the heart and multifractality in human heart beat dynamics, nonlinear dynamics theory has become a more and more important tool to analyze heart rate variance disease. Thinking about the difference of ECG signal Hurst exponent, we use nonlinear dynamics methods: Permutation Entropy(PE) [12], [13]Regularization Dimension(RD) [14] and Multifractal Spectrum(MS) [15], to detect the onset of ventricular arrhythmia and distinguish which one, VT, VF or VFL, the ventricular arrhythmia is. To test our algorithm, four MIT-BIH databases are used. Finally we give the calculating result and detailed discussion of this method.

2. The problem introduce

Here let us consider a time series Rtn. Let τ denote the time span of the whole discrete series. The cumulative sum of the difference between time series and their mean, X(n, τ), is defined:

$$X(n,\tau)=\sum_{u=1}^{n}[Rtn(u)-\sum_{n=1}^{\tau}Rtn(n)/\tau]$$

The range R denotes the difference between the maximum and minimum of X(n, τ), and S denotes standard deviation of the series. R and S are defined respectively as:

$$R(\tau)=\max_{n\in[1,\tau]}X(n,\tau)-\min_{n\in[1,\tau]}X(n,\tau),\qquad S(\tau)=\sqrt{\sum_{n=1}^{\tau}[Rtn(n)-\sum_{n=1}^{\tau}Rtn(n)/\tau]^2\Big/\tau}.$$

Both R and S are functions of τ. It is found that, for the statistical performance of the time series, the ratio R/S is very well described by the following empirical equation: $R/S=(c\tau)^H$, where c is a constant, and H the Hurst exponent. Figure 1. shows the typical Fractional Brownian motion(H=0.5) and its Hurst exponent return estimation.

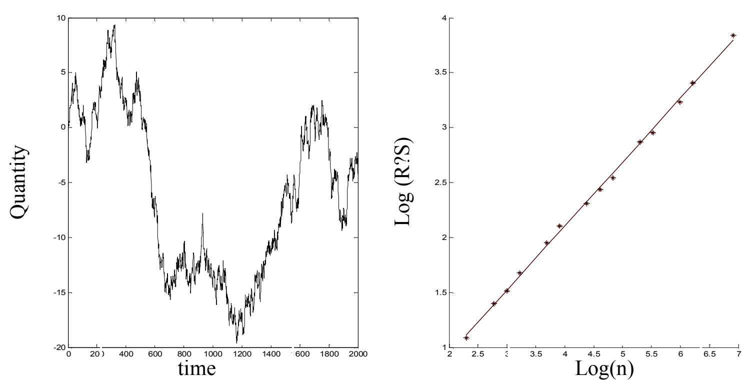

Figure 1. Fractional brownian motion(H=0.5) and its hurst exponent return estimation

Further, we calculate the Hurst exponent of SR、VT and VF signal from MIT as Figure 2. The statistic results(H_{SR}=0.8126±0.0519, H_{VA}=0.6085±0.0427) show that the Hurst parameter has some difference between SR and VA.

Hurst had pointed out that whether exponent greater or smaller than 0.5 is an important indication of the general trend in the data evolution. When the exponent is 0.5, which is called random walk, otherwise deviating random walk. A Hurst exponent greater than 0.5 ($0.5<H<1$) indicates a persistent behavior. It means that if the curve has been increasing for a while, it is likely to increase for another period, and if the curve has been decreasing, it is probably to continue decreasing. Contrarily, a Hurst exponent less than 0.5 ($0<H<0.5$) indicates an anti-persistent behavior. That is to say, after a period of decreasing, the data tend to increase in the next period, and vice versa. In the extreme when $H\rightarrow 0$, the data trend changes totally irregularly and therefore becomes completely unpredictable, exhibiting the behavior of a white noise. In the other extreme, H=1 means that the series form a straight line with a non-zero slope, and its future values are completely predictable.

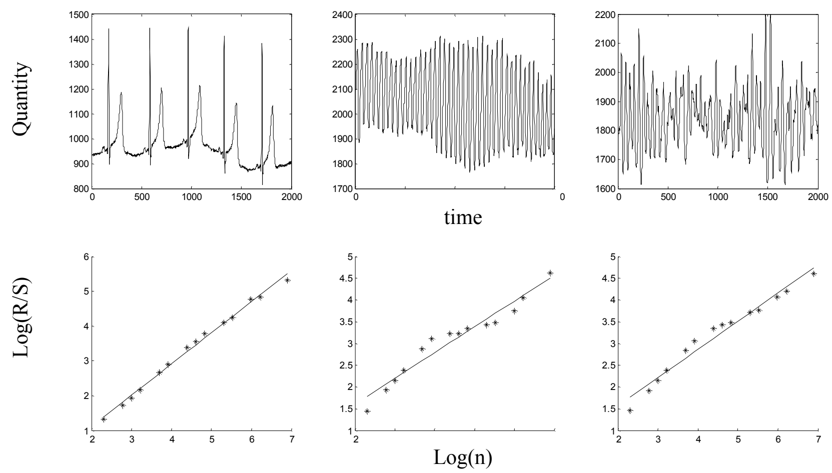

Figure 2. ECG signal and their hurst exponent SR(left), VT(middle), VF(right)

A simple relation H+D=2 exists where D is the fractal dimension. These parameters may be used to indicate the hidden statistical properties of a random process. So we can calculate the Hurst dimension easily($D_{SR}\approx 1.23$, $D_{VA}\approx 1.64$). The result indicates that through the (multi)fractal analysis we may be get more information about ECG signal.

3. Theory and methods

Here, we take ECG analysis into Fractal World and achieve very important result. First, let's begin with nonlinear theory.

3.1. Permutation Entropy

Permutation Entropy(PE) was introduced by Bandt and Pompe [12] as a convenient means of evaluating complexity parameters for time series based on comparison of neighboring values. For some well-known chaotic dynamical systems it is shown that PE behaves similar to Lyapunov exponents. It is a simple, fast, robustness and invariant method with respect to nonlinear monotonous transformations. To illustrate the method, let us first embed a scalar time series {x(i), i=1,2, . . . } to a m-dimensional space with delay time l: Xi=[x(i), x(i+l) , . . . , x(i+(m−1)l)], where m is the embedding dimension and l the delay time. For a arbitrary i, the m number of real values Xi=[x(i), x(i+l) , . . . , x(i+(m−1)l)] can be arranged in an increasing order: [$x(i+(j_1-1)l)\leq x(i+(j_2-1)l)\leq . . . \leq x(i+(j_m-1)l)$]. When an equality occurs, e.g., $x[i+(j_{i1}-1)l]=x[i+(j_{i2}-1)l]$, we order the quantities x according to the values of their corresponding j's, namely if $j_{i1}<j_{i2}$, we write $x(i+(j_{i1}-1)l)\leq x(i+(j_{i2}-1)l)$. Hence, any vector Xi is uniquely mapped onto (j_1 , j_2 , . . . , j_m), which is one of the m! permutations of m distinct symbols (1,2, . . . ,m). It is clear that each point in the m-dimensional embedding space, indexed by i, can be mapped onto one of the m! permutations. When each such permutation is considered as a symbol, then the reconstructed trajectory in the m-dimensional space is represented by a symbol sequence. The number of distinct symbols can be at most m!. Let the probability distribution for the distinct symbols be P_1 ,P_2 , . . . ,P_K, where $K\leq$ m!. Then the PE

for the time series {x(i) , i=1,2, . . . } is defined as the Shannon entropy for the K distinct symbols $H_p(m) = -\sum_{j=1}^{k} P_j \ln P_j$.

When P_j=1/m!, then $H_P(m)$ attains the maximum value ln(m!). For convenience, we always normalize $H_P(m)$ by ln(m!), and denote: $0 \leq H_P = H_P(m)/\ln(m!) \leq 1$. Thus H_p gives a measure of the departure of the time series under study from a complete random one: the smaller the value of H_p, the more regular the time series is [13]. Figure 3 shows a PE calculation example of logistic map [17].

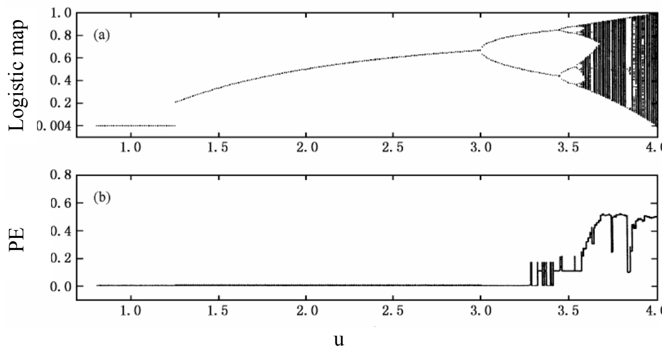

Figure 3. Logistic map (a) and its permutation entropy(b)

3.2. Regularization Dimension

Regularization Dimension(RD) [14], [18] is defined as the following way: One computes smoother and smoother versions of the original signal, obtained simply through convolution with a kernel. If the original signal is "fractal", its graph has infinite length, while all regularized versions have finite length. When the smoothing parameter tends to 0, the smoothed version tends to the original signal, and its length will tend to infinity. The regularization dimension measures the speed at which this convergence to infinity takes place. In many cases, this will coincide with the usual box dimension. In general, it can be shown that the regularization dimension is more precise than the box dimension, in the sense that it is always smaller, but still larger than the Hausdorff dimension. In addition, the regularization dimension lends to more robust estimation procedures for various reasons. One of them is that we may choose the regularization kernel. Also, the smoothed versions are adaptive by construction. Finally, the smoothing parameter can be varied in very small steps, as box sizes have to undergo sudden changes.

Given the signal S=s(t) as a function of time, consider a kernel function X=x(t) , such that $\int X = 1$. We define $X_a(t) = (1/a)x(t/a)$ as the dilated version of X at scale a; then $S_a = S \bullet X_a$ can be defined as the convolution of S with X_a. The length of the signal S on a finite interval T is given by $L_a = \int_T \sqrt{1 + S_a'(t)^2}\, dt$. According to the above assumptions and definitions, the regularization dimension of a signal S is defined as: $\dim_R(S) = 1 + \lim_{a \to 0} \frac{\log(L_a)}{-\log(a)}$.

3.3. Multifractal Spectrum

The Multifractal Spectrum(MS) is the singularity spectrum f(*a*) (fractal dimension) correlated with a (Hölder exponent). Chhabra and Jensen [15] developed a simple and precise method for the direct calculation of the multifractal spectrum from experimental data. We can consider a normalized time series as a singular measure P, calculating the f(*a*) curve covering the measure with boxes of length L and computing the probabilities P*i*(L) in each box. First a 1-parameter manifold of normalized measures $\mu_i(q)$ is defined:

$$\mu_i(q,L) = [P_i(L)]^q \Big/ \sum_j [P_j(L)]^q$$

with q the qth moment of the measure. The fractal dimension can be obtained from

$$f(q) = \lim_{L \to 0} \frac{\sum_i \mu_i(q,L) \log[\mu_i(q,L)]}{\log L}$$

and the mean strength of the singularity is obtained from

$$a(q) = \lim_{L \to 0} \frac{\sum_i \mu_i(q,L) \log[P_i(L)]}{\log L}$$

These equations provide a relationship between a Hausdorff dimension f and the average singularity strength a as implicit functions of the parameter *q*. According with this interpretation a multifractal can be visualized as an interwoven ensemble of independent monofractals of dimension f(a_i). On each of these fractals i the observable P scales with the Hölder exponent *a*.

3.4. Methods

We use aforementioned methods to analyze ventricular arrhythmia: 1)With proper embedding dimension and delay time, PE can easily detect the onset of ventricular arrhythmia, according to the change of ECG complexity; 2)In electrocardiography, we know that VF is chaotic and disordered, VT appears regular but not smooth and VFL is regular and smooth. So we use RD to analyze normal ECG and ventricular arrhythmia and classify them; 3)MS has been already used in the research of ECG and MS area has its significance on VT/VF analysis [19]. Finally, we compute MS area to distinguish VT/VF.

In order to get PE of ECG signal, we calculate the reconstruction delay time(*l*) and embedding dimension(m) according to GP algorithm [20]. Then we choose a slip window of 2000 points to compute PE, so that the slip window should not be small to keep the conclusive statistics and not be too large to obtain transitional signals accurately. To improve computation efficiency, the window slips 100 points each time. And we find that every ECG signal has its permutation entropy, when ventricular arrhythmia happens PE always decreases. But we can't use a fixed value to distinguish the onset of ventricular arrhythmia, so we compute 1th order polynomial approximation of PE every 10 seconds and take the slope as detection criterion.

We selected a Gaussian kernel and 64 progressive dilations of the kernel for the regularization process. A least square regression is used to fit the log-log plot in order to calculate the actual RD. So, SR, VFL and VT/VF can easily be confirmed. Then we calculate the MS of VT/VF after the onset of ventricular arrhythmia, from the MS area we differentiate VT or VF at last.

4. Materials and results

In this paper, four databases are used to test our algorithm. They are MIT-BIH Normal Sinus Rhythm Database, MIT-BIH Arrhythmia Database, Creighton University Ventricular Tachyarrhythmia Database and MIT-BIH Malignant Ventricular Arrhythmia Database.

First, we select 60 signals with the length of 50 seconds, which transform from SR to ventricular arrhythmia. Reconstruction delay(l) and embedding dimension(m) are computed. And we choose m=4 and l=5. Then we calculate the PE and its 1th order polynomial approximation, as shown in figure 4 and figure 5. When taking -0.01 as the minimum slope of SR, The detection accuracy is 93.33%.

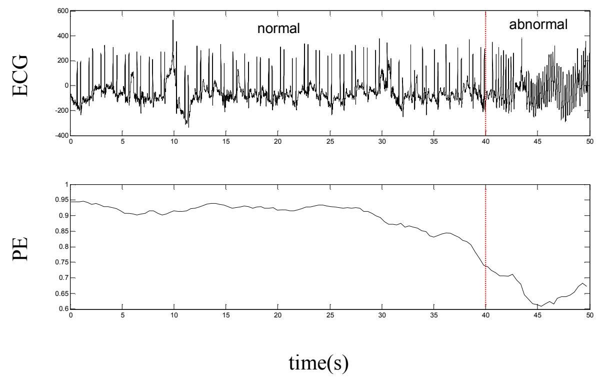

Figure. 4 ECG signal with VA and its permutation entropy.

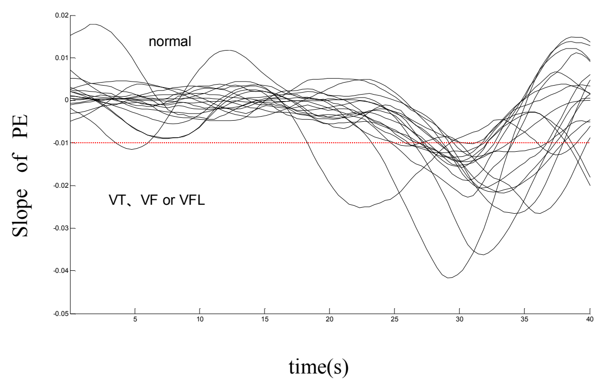

Figure. 5 The slope of 1th order polynomial approximation of permutation entropy every 10 seconds.

Then we select 3 groups and each group has 50 signals with the length of 4 seconds to calculate their RD. When we use 1.16 dimension and 1.35 dimension to classify SR, VFL and VT/VF, we get the accuracy is 100%, as shown in figure 6 and figure 7.

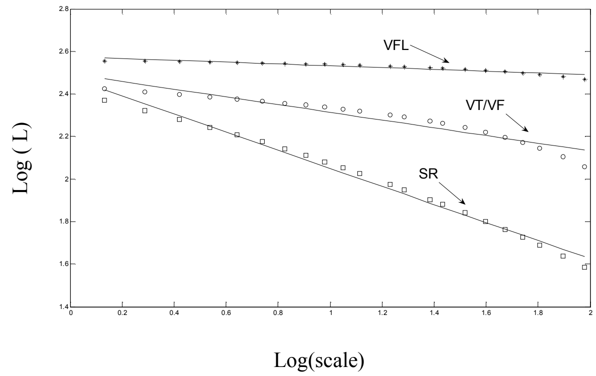

Figure 6. The regularization dimension of SR and VA

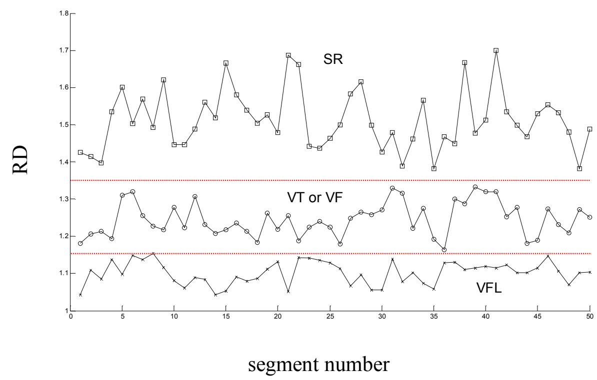

Figure 7. Classification between SR and VA with regularization dimension

Finally, according to Ref. [19], we calculate MS area to distinguish VT and VF. We select 2 groups and each group has 50 signals with the length of 5 seconds and the area threshold 0.005. When MS area is small than the threshold, the signal is VT, or it is VF. The calculation accuracy is 98%, as shown in figure 8.

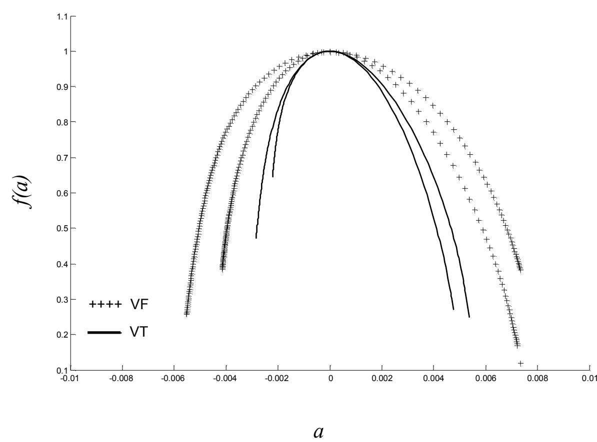

Figure 8. Multifractal spectrum of VF and VT

As a whole, the accuracy of detecting onset of ventricular arrhythmia and confirming which ventricular arrhythmia is, is VFL 93.33%, VT and VF 91.47%.

5. Discussion and conclusion

In this paper, we first talk about the fractal random walk characteristics of SR and VA. From the calculation result, we can see the persistent behavior of SR is stronger than VA, it indicates that VA is more approaching to the random behavior. And VA is a very complex series, which may be chaos or noise, as can be seen from PE, but it need more time to study. Then, we use three methods of nonlinear dynamics to detect the onset of ventricular arrhythmia and distinguish VT, VF or VFL. These theory explains the intrinsic character of ECG signal, not like other time−frequency analysis methods, it gets better detection effect. Also, PE and RD are easily programmed to achieve, and the computing processing is very quickly. Only MS computation behaves not so well. Here, we need to discuss PE, a very simple entropy algorithm. It has a good ability to analyze complexity of time series in statistics. From figure 4 and figure 5, we can see when ventricular arrhythmia happens, PE will decrease. Using polynomial approximation can find trend of PE, but it seems not very precise. So, we need to find a better way to get better result according to PE. At last, some interpretation should be put forward. Hurst estimation is an empirical and statistical method, and Hurst dimension is not the fractal dimension, such as information、box、regularization、correlation or Rényi spread dimension. So the calculating results have some difference.

6. References

[1] A.B. Smith, C.D. Jones, and E.F. Roberts, “Article Title”, Journal, Publisher, Location, Date, pp. 1-10.

[2] Jones, C.D., A.B. Smith, and E.F. Roberts, Book Title, Publisher, Location, Date.

[1] S. R. Gauna, A. Lazkano, J. Ruiz, and E. Aramendi, “Discrimination between Ventricular Tachycardia and Ventricular Fibrillation Using the Continuous Wavelet Transform”, IEEE Computers in Cardiology, 2004, pp. 21–24.

[2] N. V. Thakor, A. Natarajan, and G. Tomaselli, “Multiway sequential hypothesis testing for tachyarrhythmia discrimination”, IEEE Trans. Biomed. Eng., 1994, pp. 480–487.

[3] S. W. Chen, P. M. Clarkson, and Q. Fan, “A robust sequential detection algorithm for cardiac arrhythmia classification”, IEEE Trans. Biomed. Eng., 1996, pp. 1120–1125.

[4] V. X. Afonso, W. J. Tompkins, “Detecting ventricular fibrillation: Selecting the appropriate time-frequency analysis tool for the application”, IEEE Eng. in Med. and Biol., 1995, pp. 152–159.

[5] S. C. Schuckers, P. Raphisak, “Distinction of Arrhythmias with the Use of Approximate Entropy”, IEEE Computers in Cardiology, 1999, pp. 347–350.

[6] X. S. Zhang, Y. S. Zhu, N. V. Thakor, and Z. Z. Wang, “Detecting Ventricular Tachycardia and Fibrillation by Complexity Measure”, IEEE Trans. Biomed. Eng., 1999, pp. 548–555.

[7] Hurst H. E., Black R. P., Simaika Y. M., Long-Term Storage: An Experimental Study, Constable, London, 1965.

[8] X. T. Wang, W. Y. Qiu, and E. Y. Ren, “Option pricing of fractional version of the Black-Scholes model with Hurst exponent H being in(1/3, 1/2)”, Chaos, Solitons & Fractals, 2001, pp. 599-608.

[9] Peters E E., “Chaos and Order in the Capital Market Second Edition”, John Wiley & Sons, New York, 1996.

[10] A. L. Goldberger, “Fractal Mechanisms in the Electrophysiology of the Heart”, IEEE Eng. Med. Bio., 1992, pp. 47-52.

[11] P. C. Ivanov, et al, “Multifractality in human heart beat dynamics”, Nature, 1999, pp. 461–465.

[12] C. Bandt, B. Pompe, “Permutation Entropy: A Natural Complexity Measure for Time Series”, Physical Review Letters, 2002, PhysRevLett.88.174102.

[13] Y. Cao, W. Tung, J. B. Gao, V. A. Protopopescu, and L. M. Hively, “Detecting dynamical changes in time series using the permutation entropy”, Physical Review E, 2004, PhysRevLett.70.046217.

[14] R. Francois, L. V. Jacques, “A Regularization Approach to Fractional Dimension Estimation”, Malta: Proceeding of Fractals 98, Malta, 1998.

[15] A. Chhabra, R. V. Jensen, “Direct determination of the f(alpha) singularity spectrum”, Physical Review Letters, 1989, pp. 1327–1330.

[16] “Massachusetts Institute of Technology, PhysioBank”, Available: http : //www.physionet.org/physiobank/ database/.

[17] W. Hou, G. Feng, W. Dong, and J. Li, “A technique for distinguishing dynamical species in the temperature time series of north China”, ACTA PHYSICA SINICA, 2006, pp. 2663-2668.
[18] D. Kontos, V. Megalooikonomou, A. Javadi, P. Bakic, and A. Maidment, “Classification of Galactograms Using Fractal Properties of the Breast Ductal Network", Biomedical Imaging: Macro to Nano, the 3rd IEEE International Symposium, 2006, pp. 1324-1327.
[19] H. Xie, H. Huang, Z. Wang and J. Yuan, “Multifractal Analysis of Ventricular Fibrillation and Ventricular Tachycardia”, Journal of Data Acquisition and Processing, 2006, pp. 69-73.
[20] M. T. Rosenstein, J. J. Collins, and C. J, De Luca, “A practical method for calculating largest lyapunov exponents from small data sets”, Physical Review D, 1993, pp. 117-134.

Authors

Daoming Zhang was born in Xuzhou, China, in 1979. Doctor of College of Information and Electrical Engineering of China University of Mining and Technology. His research interests: nonlinear dynamics、signal processing and Pattern Recognition.

Guojun Tan was born in Miquan, China, in 1962. Professor of College of Information and Electrical Engineering of China University of Mining and Technology. Now, he is in America as a visit scholar of North Carolina State University. His research interests: automatic control、power electronics and electrical drive.

Jifei Hao was born in Shulan, China, in 1955. Professor of College of Information and Electrical Engineering of China University of Mining and Technology. His research interests: control theory and intelligent instruments.