Computers and Chemical Engineering 23 (1999) 899–906

Computers & Chemical Engineering

Application of wavelets and neural networks to diagnostic system development, 1, feature extraction

B.H. Chen, X.Z. Wang *, S.H. Yang, C. McGreavy

Department of Chemical Engineering, The University of Leeds, Leeds LS2 9JT, UK

Received 14 July 1997; received in revised form 9 March 1999; accepted 9 March 1999

Abstract

An integrated framework for process monitoring and diagnosis is presented which combines wavelets for feature extraction from dynamic transient signals and an unsupervised neural network for identification of operational states. Multiscale wavelet analysis is used to determine the singularities of transient signals which represent the features characterising the transients. This simultaneously reduces the dimensionality of the data and removes noise components. A modified version of the adaptive resonance theory is developed, which is designated ARTnet and uses wavelet feature extraction as the substitute of the data pre-processing unit. ARTnet is proved to be more effective in dealing with noise contained in the transient signals while retains being an unsupervised and recursive clustering approach. The work is reported in two parts. The first part is focused on feature extraction using wavelets. The second part describes ARTnet and its application to a case study of a refinery fluid catalytic cracking process.

1. Introduction

In modern process plants controlled by distributed control systems, the role of operators has changed from being primarily concerned with control to a broader supervisory responsibility: analysing operational data, identifying unusual conditions as they develop and responding rapidly and effectively by taking corrective actions. This is a challenging task because of the overwhelming volume of data operators have to deal with. In recent years there has been a significant progress in applying intelligent systems for process monitoring and diagnosis. This includes the use of neural networks, multivariate statistical analysis, expert systems as well as qualitative simulation. It is recognised that in process monitoring and diagnosis, dynamic trend signals are often more important than variable values at the current sampling instant and that therefore a critical issue is how to discriminate dynamic transients automatically in computer based systems. Computer based processing of dynamic trend signals is aimed at noise removal and dimension reduction using minimum data points to capture the features characterising the trend signals. Various approaches have been proposed and their effectiveness depends largely on how the processed information is to be used, i.e. by human experts, expert systems or neural networks. In this work, an integrated framework, ARTnet is developed and subsequently applied to a case study of a refinery fluid catalytic cracking process. ARTnet is a modified version of the adaptive resonance theory (ART2) (Carpenter and Grossberg, 1987; Whiteley & Davis, 1992, 1994; Whiteley, Davis, Mehrotra, & Ahalt, 1996) which uses wavelet transforms as the substitute of the data preprocessing unit of ART2.

The work is reported in two parts. The first part is focused on feature extraction from dynamic transient signals using wavelet transforms and the second part is concerned with the introduction of ARTnet and its application to a case study of a refinery fluid catalytic cracking process. The first part is organised as follows. In Section 2 some representative approaches for feature extraction are briefly reviewed. This naturally leads to the introduction of wavelet multiscale analysis for feature extraction in Section 3. Wavelet multiscale analysis finds the extrema of a transient signal and an important

* Corresponding author. Tel.: +44-113-233-2427; fax: +44-113-233-2405.

E-mail address: x.z.wang@leeds.ac.uk (X.Z. Wang)

0098-1354/99/$ - see front matter
PII: S0098-1354(99)00258-6

issue is how to remove the effects of noise components and achieve consistent results in different scales. This is the subject of Section 4.

2. Previous work on feature extraction of dynamic transients

This section briefly reviews some of the previous work on feature extraction. Feature extraction is basically a transformation of the data composing a dynamic trend to a lower dimensionality. An important property of such a transformation is that it is information preserving, that is, data is reduced by removing redundant components while preserving, in some optimal sense, information which is crucial for pattern discrimination.

Some researchers have adapted the episode representation technique originated by William (1986) to qualitative interpretation of transient signals. Janusz and Venkatasubramanian (1991) developed an episode approach that uses nine primitives to represent any plots of a function. Each primitive consists of the signs and the first and second derivatives of the function. Therefore, each primitive posses the information about whether the function is positive or negative, increasing, decreasing, or not changing and the concavity. An episode is an interval described by only one primitive and the time interval the episode spans. A trend is a series of episodes that when grouped together can completely describe the dynamic feature. The approach automatically converts on-line sensor data to qualitative classification trees. Cheung and Stephanopoulos (1990) developed a slightly different approach called triangular-episode that uses seven triangle components to describe a dynamic trend. Bakshi and Stephanopoulos (1994, 1996) used wavelet decomposition of functions in different scales and zero-crossing of wavelet derivatives to find the inflections of decomposition. In this way, episodes can be identified automatically by computers. Based on episode analysis, dynamic trends can be interpreted as symbolic representations. The main idea of dynamic trend interpretation using episode approaches is to classify a trend such as increasing or decreasing pieces. This interpretation is some times not enough and inadequate in process analysis. Furthermore, there is no noise filtering in any of the episode based approaches, which significantly limits the trend representation and identification capability.

Whiteley and Davis (1992) applied back-propagation neural networks (BPNN) to convert numerical sensor data into symbolic abstractions. The major limitation of this approach is that it requires training data to train the model first.

The most well known technique for signal analysis is probably the Fourier transform and it is therefore necessary to mentioned it here. Fourier transform uses sine and cosine as its building blocks to decompose a function into a sum of frequency components. However, Fourier transform does not show how frequency varies with time, therefore it is not able to detect when a particular event took place. It means that the non-stationary feature of the signal is not captured. The short-time Fourier transform is able to overcome this limitation by sliding a window over the signal in time. However in time-frequency analysis of a non-stationary signal, there are two conflicting requirements. The window width must be long enough to give the desired frequency resolution but must also be short enough to lose track of time dependent events. While it is possible to optimise the design of window shapes to optimise, or trade-off time and frequency resolution, there is a fundamental limitation on what can be achieved, for a given fixed window width (Dai, Joseph & Motard, 1994).

3. Feature extraction using wavelet transform

A very brief introduction of wavelet transformation for signal processing is now presented. Then the method employed in this study for feature extraction using wavelets is introduced and illustrated using examples.

3.1. Signal transformation using wavelets

Wavelet transformation is designed to address the problem of non-stationary signals. It involves representing a time function in terms of simple, fixed building blocks, termed wavelets. These building blocks are actually a family of functions which are derived from a single generating function called the mother wavelet by translation and dilation operations. Dilation, also known as scaling, compresses or stretches the mother wavelet and translation shifts it along the time axis.

The mother wavelet satisfies

$$\int_{-\infty}^{+\infty} \psi(t)\, \mathrm{d}t = 0 \tag{1}$$

and the translation and scaling operations on $\psi(t)$ creates a family of functions,

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\, \psi\left(\frac{t-b}{a}\right) \tag{2}$$

The parameter a is a scaling factor and stretches (or compresses) the mother wavelet. The parameter b is a translation along the time axis and simply shifts a wavelet and so delays or advances the time at which it is activated. Mathematically delaying a function $f(t)$ by t_d is represented by $f(t-t_d)$. The factor $1/\sqrt{a}$ is used to ensure the energy of the scaled and translated versions are the same as the mother wavelet.

The stretched and compressed wavelets through scaling operation are used to capture the different frequency components of the function being analysed. The translation operation, on the other hand, involves shifting of the mother wavelet along the time axis to capture the time information of the function to be analysed at a different position. In this way, a family of scaled and translated wavelets can be created using scaling and translation parameters a and b. This allows signals occurring at different times and having different frequencies to be analysed. In contrast to the short-time Fourier transform, which uses a single analysis window function, the wavelet transform can use short windows at high frequencies or long windows at low frequencies. Thus wavelet transform is capable of zooming-in on short-lived high frequency phenomena and zooming-out on sustained low frequency phenomena. This is the main advantage of the wavelet over the short-time Fourier transform.

Wavelet transform can be categorised into continuous and discrete. Continuous, in the context of wavelet transform, implies that the scaling and translation parameters a and b change continuously. However, calculating wavelet coefficients for every possible scale can represent a considerable effort and result in a vast amount of data. Therefore discrete parameter wavelet transform is often used. The discrete parameter wavelet transform uses scale and position values based on powers of two-so-called dyadic scales and positions and makes the analysis much more efficient, whilst remaining accurate. To do this, the scale and time parameters are discretised as follows,

$$a = a_0^m, \qquad b = nb_0a_0^n \quad m,\ n \text{ are integers} \tag{3}$$

The family of wavelets $\{\psi_{m,n}(t)\}$ is given by

$$\psi_{m,n}(t) = a_0^{-m/2}\psi(a_0^{-m}t - nb_0) \tag{4}$$

resulting in a discrete wavelet transform (DWT) having the form

$$\mathrm{DWT}_f(m, n) = \langle f,\ \psi_{m,n}\rangle = a_0^{-m/2}\int_{-\infty}^{+\infty} f(t)\psi(a_0^{-m}t - nb_0) \tag{5}$$

Mallat (1989) developed an approach for implementing this using filters. For many signals, the low frequency content is the most important part. The high frequency content, on the other hand provides flavour or nuance. In wavelet analysis the low frequency content is called the *approximation* and the high frequency content is called the *detail*. The filtering process uses *lowpass* and *highpass* filters to decompose an original signal into the *approximation* and *detail* parts. It is not necessary to preserve all the outputs from the filters. Normally they are *downsampled* and keep only the even components of the *lowpass* and *highpass* filter outputs. The decomposition can be iterated, with successive approximations being decomposed in turn, so that one signal is broken into many lower-resolution components.

In the case of a discrete wavelet transform, reconstruction of the original signal is not guaranteed. Daubechies (1992) developed conditions under which the $\{\psi_{m,n}\}$ form an orthonomal basis. Usually, $a_0 = 2$ and $b_0 = 1$ are used, although any values can be used. In this case, both the transform and reconstruction are complete because the family of wavelets form an orthonormal basis.

3.2. Singularity detection using wavelets for feature extraction

Singularities often carry the most important information in signals. Singularities of a signal can be used as the compact representation, i.e. the features of the original signal. Mathematically, the local singularity of a function is measured by Lipschitz exponents (Mallat & Hwang, 1992). Mallat and Hwang (1992) proved that the local maxima of the wavelet transform modulus detects the locations of irregular structures and provides numerical procedures for computing the Lipschitz exponents. Within the framework of scale-space filtering, inflexion points of $f(t)$ appear as extrema for $\partial f(t)/\partial t$ and zero crossing for $\partial^2 f(t)/\partial t^2$, so Mallat and Zhong (1992) suggests using a wavelet which is the first derivative of a scaling function $\Phi(t)$,

$$\psi(t) = \frac{\mathrm{d}\phi(t)}{\mathrm{d}t}$$

with a cubic spine being used for the scaling function. Bakshi and Stephanopoulos (1996) used the inflexion points as the connection points of episode segments of a signal.

The wavelet modulus maxima and zero-crossing representations were developed from underlying continuous-time theory. For computer implementation, this has to be cast in discrete-time domain. Berman and Baras (1993) proved that wavelet transform extrema/zero-crossing provide stable representations of finite length discrete-time signals. A more complete discrete-time framework for the representation of the wavelet transform was developed by Cvetkovic and Vetterli (1995) and therefore is used in this study. They designed a non-subsampled multi-resolution analysis filter bank to implement the wavelet transform. Using this filter bank, the wavelet function can be selected from a wider range than the B-spline in Mallat's method.

Non-subsampled multi-resolution analysis was used to determine singularities of a signal. An octave band non-subsampled filter bank with analysis filters $H_0(z)$ and $H_1(z)$ is shown in Fig. 1. In this method, a wavelet transform refers to the bounded linear operators

$W_j: l^2(Z) \rightarrow l^2(Z)$; $j = 1, 2, \ldots j+1$. The operators W_j, are the convolution operators with the impulse responses of the filters:

$$V_1(z) = H_1(z)$$

$$V_2(z) = H_0(z) H_1(z^2)$$

$$V_j(z) = H_0(z) \cdots H_0(z^{2^{j-2}}) H_1(z^{2^{j-1}})$$

$$V_{j+1}(z) = H_0(z) \cdots H_0(z^{2^{j-2}}) H_0(z^{2^{j-1}})$$

The multiresolution procedure depicted in Fig. 1 can be described less rigorously. Fig. 1 shows four steps, or four scales. In the first scale, the original signal is split into *approximation* $A_x{}^1$ and *detail* $D_x{}^1$. The *detail* $D_x{}^1$ is supposed to be mainly the noise components of the original signal. $A_x{}^1$ is further decomposed into *approximation* $A_x{}^2$ and *detail* $D_x{}^2$, $A_x{}^2$ to $A_x{}^3$ and $D_x{}^3$ and $A_x{}^3$ to $A_x{}^4$ and $D_x{}^4$. In each step the extrema of the *detail* are found. Apparently, in the first few steps, the extrema are both as a result of the noise and the trend of the noise-free signal. With scales being increased, the noise extrema will gradually be removed while the extrema of the noise-free signal remain. In this way, through multi-scale analysis and extrema determination, the extrema of the noise-free signal can be found, which are regarded as the features of the signal.

For the representation of extrema, it is convenient to use a finite impulse response (FIR) wavelet filter. The FIR is a filter with the sequence $\{\alpha_k : k \in Z\}$ and has only K non-zero terms. A typical example is the Haar wavelet, which has only two non-zero coefficients. Daubechies' wavelets (Daubechies, 1992) are also FIR filters and smoother than the Haar wavelet. Daubechies' wavelets having more coefficients are smoother and have higher vanishing moments. They also require less computational effort as they are constructed by filter convolution.

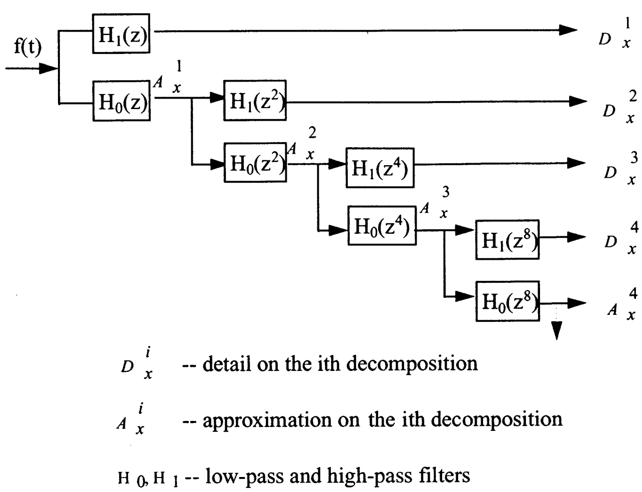

Fig. 1. An octave band non-subsampled filter bank.

The Daubechies' scale and wavelet functions are expressed as

$$\phi(t) = \sum_k h(k) \phi(2t - k) \tag{6}$$

$$\psi(t) = \sum_k g(k) \phi(2t - k) \tag{7}$$

where $\{h(k)\}$ is the low-pass filter coefficients and $\{g(k)\}$ the band-pass filter coefficients.

Daubechies' wavelets have a maximum number of vanishing moments for the support space. The vanishing moments of the wavelets also have a different number of coefficients. Using wavelets with more vanishing moments has the advantage of being able to measure the Lipschitz regularity up to a higher order, which is helpful in filtering noise, but it also increases the number of maxima lines. The number of maxima for a given scale often increases linearly with the number of moments of the wavelet. In order to minimise computational effort, it is necessary to have a minimum number of maxima to detect the significant irregular behaviour of a signal. This means choosing a wavelet with as few vanishing moments as possible but with enough moments to detect the Lipschitz exponents of the highest order components of interest.

In this study, an eight coefficient 'least-asymmetric' Daubechies' wavelet is used as a filter. The scale and wavelet function for this filter are illustrated in Fig. 2.

A signal $f(t) = \sin(t)$ and its extrema of wavelet analysis using non-subsampled filter bank with Daubechies' eight coefficients least asymmetry wavelet is illustrated in Fig. 3, which shows that extrema of wavelet analysis correspond to the singularities of the signal. In Fig. 3b, the wavelet is used as filter and the first singularity of the signal in Fig. 3a corresponds to minimum of wavelet analysis. In Fig. 4 it is a maximum because a different wavelet is employed. The former is used here.

3.3. Noise extrema removal

The extrema obtained from wavelet multi-resolution analysis correspond to the singularities of the signal, which may also include those produced by noise, depending on the analysis scales. Therefore, in feature extraction it is necessary to further identify and filter out noise extrema from wavelet transform. The most classical technique of removing noise from a signal is to filter it. Part of the noise is removed but it may also smooth the signal singularities at the same time. Mallat and Hwang (1992) and Mallat and Zhong (1992) developed a technique for evaluating noise extrema in wavelet analysis. They found that some noise maxima increase on average when the scale decreases or don't propagate to larger scales. These are the modulus maxima which are mostly influenced by noise fluctuations.

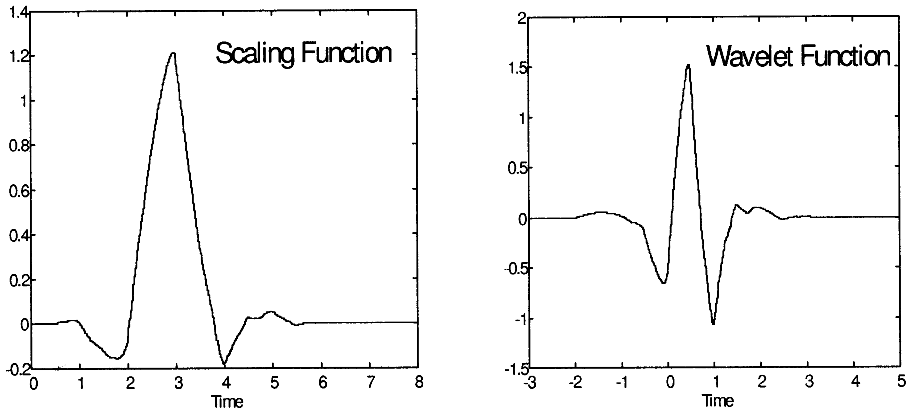

Fig. 2. The 'least-Asymmetric' scale function and wavelet function.

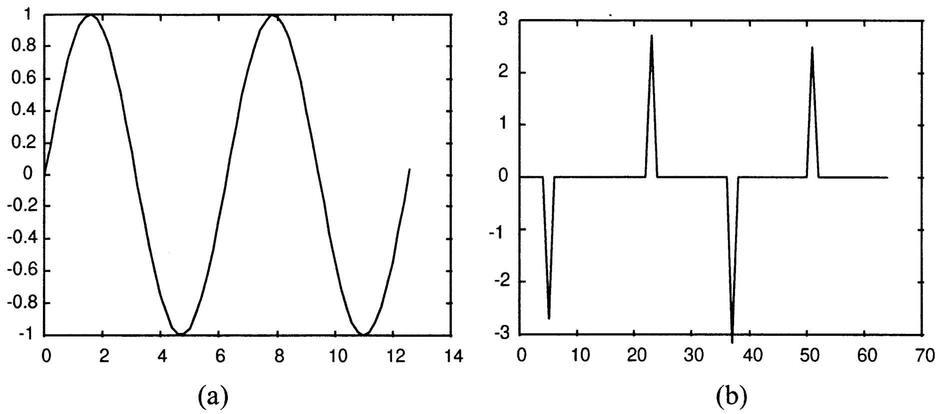

Fig. 3. Signal (a) and its extrema (b) of wavelet analysis using Daubechies' eight coefficients wavelet.

Fig. 5 and Fig. 6 illustrate this idea. In Fig. 5, three different noise frequencies are studied. The wavelet multi-resolution analysis is shown on the left and extrema of wavelet analysis are on the right. Clearly, the extrema will decrease and then disappear as the scale increases.

Fig. 6 shows a signal which is basically the sine in Fig. 3a corrupted by white noise as well as the wavelet multi-scale extrema analysis. Noise components are reduced and then disappear as the scale increases. The results for scales-4 and -5 are similar to that of Fig. 3b which is noise-free. This shows that the extrema of the trend signal are retained while noise extrema are filtered.

3.4. Piece-wise processing

Two observations are made about the above discussions. Firstly, extrema analysis using wavelet multiresolution analysis remains steady with the increase of scales, so the representation is steady. For example, in Fig. 6 when the scale is increased from four to five, the four extrema remain. Secondly, the location of extrema may slightly shift with time as scale increases. In Fig. 6, the extrema representation in scale-4 is a vector of dimension 70,

Scale-4 = (… x5… x23… x37… x53…)

where x5 stands for a non-zero datum in column 5.While in scale-5, it becomes

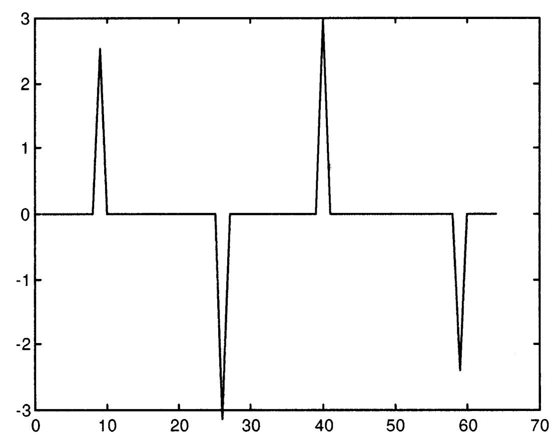

Fig. 4. Extrema of wavelet analysis using Daubechies' ten coefficients wavelet.

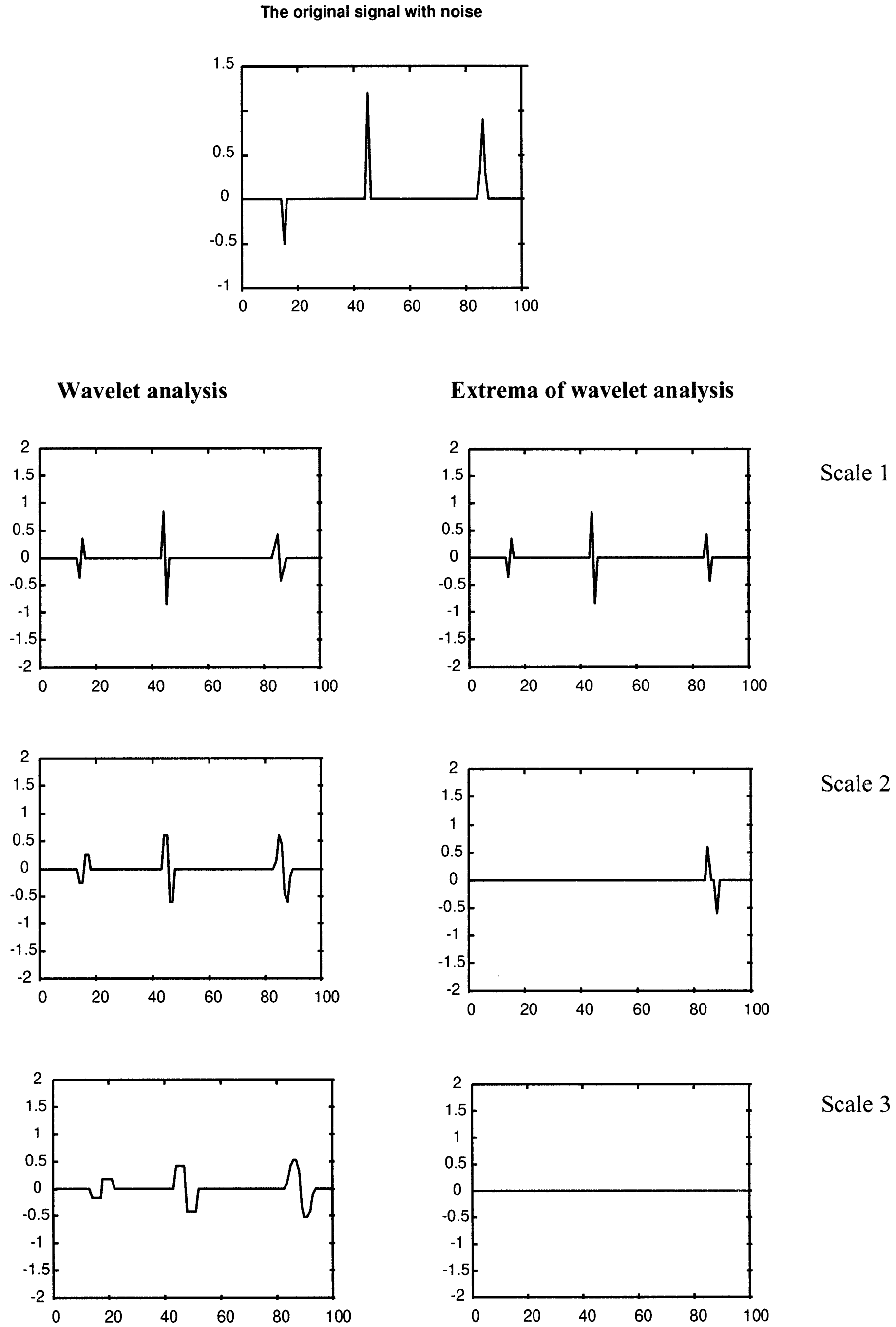

Fig. 5. Noise signal, its wavelet transformation and the extrema of wavelet analysis.

Scale-5 = (… x7… x22… x38… x54…)

It is obvious that non-zero datum in the position 5 of scale-4 is shifted to the position 7 of scale-5. This inconsistency should be avoided. For instance (2, 0 … 0, 3) and (2, 0 … 3, 0) should be considered different. This is necessary especially when the trends of a variable at different operations conditions are considered.

The extrema representation can be highly sparse vectors. This is true for process dynamic responses which are slow in frequency. The method we used is called piece-wise processing. The idea is to map a highly sparse vector to a denser vector by dimension reduction. For example, with scale-4 and scale-5 discussed above, if the piece-wise sub-region is fixed as four data points, then scale-4 and scale-5 will be transformed to vectors of dimension 18.

Scale-4′ = (… x2 … x6 … x10 … x13 …)

Scale-5′ = (… x2 … x6 … x10 … x13 …)

It is clear that after piece-wise processing, the dimension is reduced and scale-4′ and scale-5′ are consistent. Therefore using piece-wise processing technique, it can achieve consistent feature extraction as well as dimension reduction.

4. Final Remarks

Features of a process dynamic transient signal are identified as the singularities and irregularities because they contain the most important information corresponding to changes of operational states. The approach developed by Mallat and Hwang (1992) and Cvetkovic and Vetterli (1995) for determining singularities and irregularities is introduced for feature extraction of dynamic transient signals of process operations, which are the extrema of wavelet analysis. An approach for noise extrema removal and piece-wise dimension reduction are also discussed. In the second part of the paper, the use of the approach to replace the data pre-processing part of the adaptive resonance theory to develop a more efficient unsupervised and recursive learning system ARTnet and describe its application to a refinery fluid catalytic cracking process is reported.

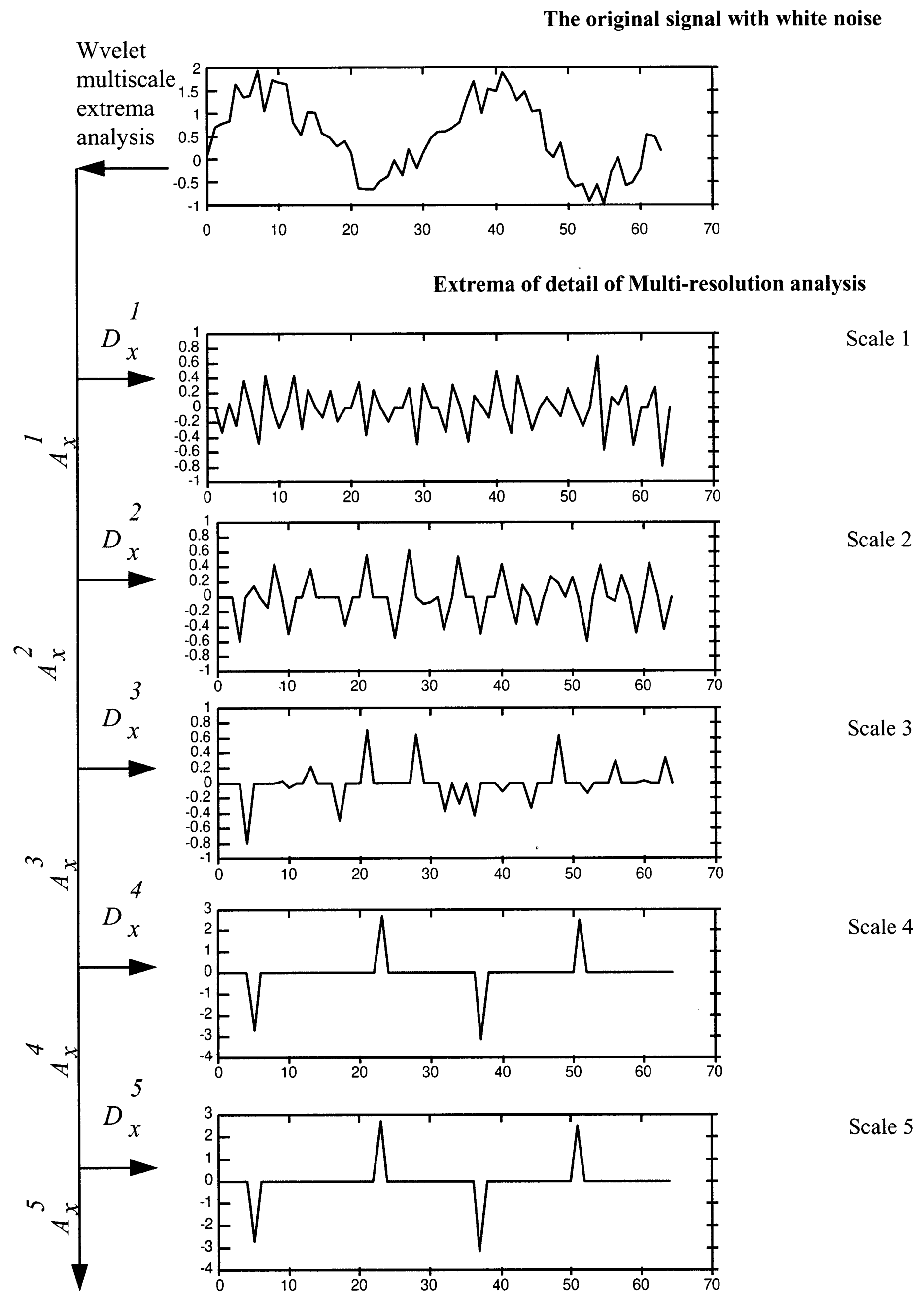

Fig. 6. Noise signal and its multi-resolution analysis. $A_x^{\,i}$, approximation of multiresolution analysis; $D_x^{\,i}$, detail.

5. Notation

A	approximation in wavelet multiresolution analysis
a	wavelet dilation parameter
a_0, b_0	discrete wavelet transform parameters
b	wavelet translation parameter
D	detail in wavelet multiresolution analysis
DWTf	discrete wavelet coefficient
$f(t)$	a function in the time domain
$g(k)$	the kth wavelet synthesis filter
H	wavelet analysis filter
$h(k)$	the kth wavelet analysis filter
m, n	discrete wavelet transform parameters
s	scale
t	time

Greek

α	Lipchitz exponent
ψ	wavelet function
$\phi(t)$	wavelet scale function or orthogonal function

Acknowledgements

The authors are indebted to the anonymous referees for their constructive comments which led to an improved manuscript. The authors acknowledge the partial financial support of the Aigis Systems, USA. The first author thanks the financial support of the Sino-British Friendship Scholarship Scheme (SBFSS).

References

Bakshi, B. R., & Stephanopoulos, G. (1996). Reasoning in time: modelling, analysis and pattern recognition of temporal process trends. In G. Stephanopoulos, & C. Han, *Intelligent Systems in Process Engineering-Paradigms from Design to Operations* (pp. 487–549).

Bakshi, B. R., & Stephanopoulos, G. (1994). Representation of process trends-3: multiscale extraction of trends from process data. *Computers and chemical Engineering*, *18*, 267–302.

Berman, Z., & Baras, J. S. (1993). Properties of the multi-scale maxima and zerocrossing representations. *IEEE Transactions on Signal Processing*, *41*, 3216–3231.

Carpenter, G. A., & Grossberg, S. (1987). ART2: self-organisation of stable category recognition codes for analogue input patterns. *Applied Optics*, *26*, 4919–4930.

Cheung, J. T. Y., & Stephanopoulos, G. (1990). Representation of process trends-1: a formal representation framework. *Computers and chemical Engineering*, *14*, 495–510.

Cvetkovic, Z., & Vetterli, M. (1995). Discrete-time wavelet extrema representation: design and consistent reconstruction. *IEEE Transactions on Signal Processing*, *43*, 681–693.

Dai, X., Joseph, B., & Motard, R. L. (1994). Introduction to wavelet transformation and time frequency analysis. In R. L. Motard, & B. Joseph, *Wavelet Applications in Chemical Engineering* (pp. 1–32). Dordrecht: Kluwer.

Daubechies, I. (1992). *Ten lectures on wavelets*. Philadelphia, PA: Society of Industrial and Applied mathematics.

Janusz, M. E., & Venkatasubramanian, V. (1991). Automatic generation of qualitative descriptions of process trends for fault detection and diagnosis. *Engineering Applications Artificial Intelligence.*, *4*, 329–339.

Mallat, S. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Pattern Analysis and Machine Intelligence*, *11*, 674–693.

Mallat, S., & Hwang, W. L. (1992). Singularity detection and processing with wavelets. *IEEE Transactions on information theory*, *38*, 617–643.

Mallat, S., & Zhong, S. (1992). Characterisation of signals from multi-scale edges. *IEEE Transaction Pattern Analysis and Machine Intelligence*, *14*(7), 710–732.

Whiteley, J. R., & Davis, J. F. (1992). Knowledge-based interpretation of sensor patterns. *Computers and Chemical Engineering*, *16*, 329–346.

Whiteley, J. R., & Davis, J. F. (1994). A similarity-based approach to interpretation of sensor data using adaptive resonance theory. *Computers and Chemical Engineering*, *18*, 637–661.

Whiteley, J. R., Davis, J. F., Mehrotra, A., & Ahalt, S. C. (1996). Observations and problems applying ART2 for dynamic sensor pattern interpretation. *IEEE Transactions on System Manual Cybernetics Part A: Systems and Humans*, *26*, 423–437.

William, B.C. (1986) Doing time: putting qualitative reasoning on firmer ground. *National Conference on Artificial Intelligence*, Philadelphia.