ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/220566501

Discrete Wavelet Transform-Based Time Series Analysis and Mining

Article *in* ACM Computing Surveys · February 2011

DOI: 10.1145/1883612.1883613 · Source: DBLP

CITATIONS
201

READS
3,663

4 authors:

Pimwadee Chaovalit
National Electronics and Computer Technology Center
31 PUBLICATIONS **1,401** CITATIONS

SEE PROFILE

Aryya Gangopadhyay
University of Maryland, Baltimore County
174 PUBLICATIONS **1,764** CITATIONS

SEE PROFILE

George Karabatis
University of Maryland, Baltimore County
68 PUBLICATIONS **1,103** CITATIONS

SEE PROFILE

Zhiyuan Chen
University of Maryland, Baltimore County
91 PUBLICATIONS **1,346** CITATIONS

SEE PROFILE

All content following this page was uploaded by Zhiyuan Chen on 09 July 2015.

The user has requested enhancement of the downloaded file.

Discrete Wavelet Transform-Based Time Series Analysis and Mining

PIMWADEE CHAOVALIT, National Science and Technology Development Agency
ARYYA GANGOPADHYAY, GEORGE KARABATIS, and ZHIYUAN CHEN, University of Maryland, Baltimore County

Time series are recorded values of an interesting phenomenon such as stock prices, household incomes, or patient heart rates over a period of time. Time series data mining focuses on discovering interesting patterns in such data. This article introduces a wavelet-based time series data analysis to interested readers. It provides a systematic survey of various analysis techniques that use discrete wavelet transformation (DWT) in time series data mining, and outlines the benefits of this approach demonstrated by previous studies performed on diverse application domains, including image classification, multimedia retrieval, and computer network anomaly detection.

Categories and Subject Descriptors: A.1 [**Introductory and Survey**]; G.3 [**Probability and Statistics**]: —*Time series analysis*; H.2.8 [**Database Management**]: Database Applications—*Data mining*; I.5.4 [**Pattern Recognition**]: Applications—*Signal processing, waveform analysis*

General Terms: Algorithms, Experimentation, Measurement, Performance

Additional Key Words and Phrases: Classification, clustering, anomaly detection, similarity search, prediction, data transformation, dimensionality reduction, noise filtering, data compression

ACM Reference Format:
Chaovalit, P., Gangopadhyay, A., Karabatis, G., and Chen, Z. 2011. Discrete wavelet transform-based time series analysis and mining. ACM Comput. Surv. 43, 2, Article 6 (January 2011), 37 pages.
DOI = 10.1145/1883612.1883613 http://doi.acm.org/10.1145/1883612.1883613

1. INTRODUCTION

A time series is a sequence of data that represent recorded values of a phenomenon over time. Time series data constitutes a large portion of the data stored in real world databases [Agrawal et al. 1993]. Time series data appear in many application domains, such as in financial, meteorological, medical, social sciences, computer networks, and business. Time series are derived from recording observations of various types of phenomena, for example, temperature, stock prices, household income, patient heart rates, number of bits transferred, product sales volume over a period of time, etc. Some complex data types, such as audio and video, are also considered time series data, since they can be measured at each point in time.

This research was supported by the Royal Thai Scholarship.
This work was conducted when P. Chaovalit was a doctoral student at the University of Maryland, Baltimore County (UMBC).
Authors' addresses: P. Chaovalit, National Science and Technology Development Agency, 111 Thailand Science Park, Pahonyothin Road, Klong 1, Klong Luang, Pathum Thani 12120, Thailand; email: pimwadee@nstda.or.th; A. Gangopadhyay, G. Karabatis, and Z. Chen, Department of Information Systems, The University of Maryland, Baltimore County (UMBC), 1000 Hilltop Circle, Baltimore, MD 21250; email: {gangopad, georgek, zhchen}@umbc.edu.
Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org.
© 2011 ACM 0360-0300/2011/01-ART6 $10.00
DOI 10.1145/1883612.1883613 http://doi.acm.org/10.1145/1883612.1883613

Time series data mining techniques analyze time series data in search of interesting patterns that were previously unknown to information users. Researchers and users perform various tasks on time series data, such as time series classification, time series clustering, rule extraction, and pattern querying. For example, when users want to gain an insight into stock prices, they explore the closing price data by clustering data into price groups. Then they may track the stocks with certain price fluctuations by performing a query. When users are familiar with the data, they may use a rule extraction technique to mine a set of rules that best govern the stock prices. To perform these interesting tasks, different techniques have already been established. One of the more recent and promising techniques is discrete wavelet transform.

Discrete wavelet transform (DWT), a technique with a mathematical origin, is very appropriate for noise filtering, data reduction, and singularity detection, which makes it a good choice for time series data processing. DWT has been around for approximately 100 years, and it has been used extensively in a wide range of areas, such as in signal processing, and specifically it is frequently employed for research in signal compression, image enhancement and noise reduction.

Time series data analysis and mining is another area where researchers have recently applied DWT techniques due to its favorable properties. Although DWT has been around for quite some time, only recently has it been adopted by database researchers to assist in data analysis and mining for time series.

DWT is a powerful tool for a time-scale multiresolution analysis on time series and has been used to break down an original time series into different components, each of which may carry meaningful signals of the original time series. Researchers have applied wide-ranging analyses on decomposition of an original time series in medical time series data, audio and video data, and image data and obtained superior results. A notable example describing the value of DWT in the decomposition of a time series comes from the medical domain: an EEG (electroencephalograph) signal is the most important measurement to assist in the diagnosis of epilepsy. In Subasi [2005], an EEG signal was broken down into several subbands using DWT, and produced better intermediate results to be fed into a classification engine. The classification engine using an artificial neural network diagnosed patients as healthy or epileptic from the decomposed subband of EEG with more than 90% accuracy when using the human experts' diagnoses as baseline. Such a system can serve suitably as a great decision support tool for medical experts.

There are many advantages in using DWT ranging from the discovery of more precise knowledge, to the development of faster mining process, all the way to the reduction of data storage requirements. In this article, we discuss and provide a strong basis for understanding the use of DWT on time series data for data analysis and mining purposes. In Section 2 we present time series data definition and characteristics. In Section 3 we present the concept of discrete wavelet transform and its multiple levels of resolution, and discuss the benefits and functionalities of DWT for time series data analysis. The functionalities include data dimensionality reduction, noise filtering, and singularity detection, which are available for multiresolution analysis. In Section 4 we discuss applications of discrete wavelet transforms in various domains of time series data analysis and mining, including (i) wavelet-based time series similarity search, (ii) wavelet-based time series classification, (iii) wavelet-based clustering, (iv) wavelet-based trend, surprise, and pattern detection, and (v) wavelet-based prediction. We conclude this article in Section 5 by summarizing the benefits of DWT, indicating research gaps, and identifying challenges involved in applying DWT to time series data analysis and mining for interested researchers.

2. TIME SERIES DATA ANALYSIS AND MINING

The growth of time series data has profoundly increased the interest in data analysis and mining of time series by both academic and industry researchers. In this article we concentrate mainly on topics relevant to wavelet-based time series data analysis and mining; nevertheless, there is a rich body of literature for generic time series data analysis and mining, which is briefly presented for comparison in Section 4, although the discussion there can by no means be considered exhaustive. For further reading on generic time series data analysis and mining, we direct the readers to examine the excellent survey articles by Keogh et al. [2004a], Keogh and Kasetty [2002], and Roddick and Spiliopoulou [1999]. We start our discussion on time series data analysis with a definition of time series. Then we introduce the characteristics of time series data.

2.1. Definition of Time Series Data

A time series is a sequence of event values which occur during a period of time. Each event occurring at each time point has a value which is recorded. The collection of all these values represents a single variable (such as an EEG signal or stock price over a time period). Therefore, a time series of a single variable contains a sequence of recorded observations of an interesting event. Formally a time series can be represented by $S = \{s_1, s_2, \ldots, s_n\}$, where S is a whole time series, s_i is the recorded value of variable s at time i, and n is the number of observations.

2.2. Time Series Data Characteristics

Time series data has some daunting characteristics for data mining: large volume, high dimensionality, hierarchy, and multivariate property. We will discuss each of these characteristics in this section.

A large volume of data in the database could pose a challenge for data analysis. With time series data mining, the situation is exacerbated even further when, for example, we use systems that constantly collect monitoring data from automatic sensors. The number of observations in a time series can often be extremely high, sometimes ranging from the order of hundreds or thousands to the order of millions or billions. The large volume of data poses a problem for data analysis and mining algorithms as larger databases take more time for data analysis and mining technique to access data and perform computations.

High dimensionality is another easily-recognized characteristic of time series data. It refers to situations when time series are long. During similarity search in time series data analysis, this leads to what is known as the *dimensionality curse* [Agrawal et al. 1993; Chan and Fu 1999; Lee et al. 2000; Man and Wong 2001]. Dimensionality curse is the situation that arises when a time series is mapped onto a k-dimensional space, where k is the number of time points. Korn et al. [1997] proposed an approach using singular value decomposition (SVD) to transform a large matrix of time series into a smaller matrix for data compression purposes, as follows. If we consider a set of time series data as having M observations, each of which has N data points, we have an M^*N matrix. The method assumes that the number M is much larger than N. With the Korn et al. [1997] technique, random accesses to data for ad hoc queries are possible with a small reconstruction error. However, Korn et al.'s [1997] approach might not be applicable to some time series datasets since this assumption may or may not hold true, depending on the length of the data. For a time series that is very long, the number N could easily exceed M [Shahabi et al. 2000]. A dataset composed of reasonably long time series with moderate number of observations may not use the approach of Korn et al. [1997].

Another characteristic of time series data is its hierarchical nature. A time series can be analyzed by its underlying time hierarchy, such as hourly, weekly, monthly, and yearly. A number of studies have investigated multilevel analysis of time series data hierarchically [Geurts 2001; Man and Wong 2001; Percival and Walden 2000; Shahabi et al. 2000]. These investigations led researchers to look for patterns by temporal semantics through the time series hierarchy. For example, Li et al. [1998] queried data from a time series database on multiple levels of abstraction. Users could also find the match for a larger sequence of events by forming together several small events. Shahabi et al. [2000] proposed a technique for analyzing trend and surprise in time series' temporal hierarchies through visualization.

The last characteristic of time series data is the multi-variate nature of some data. Time series data analysis often studies one variable, but sometimes deals with time series data consisting of multiple related variables. For example, weather data consists of well-known measurements such as temperature, dew point, humidity, etc. Even though most of the work in time series data analysis and mining has focused on time series data for one variable, studies on multiple time series have appeared in the literature [Dillard and Shmueli 2004; Huhtala et al. 1999; Shmueli 2004], where they sometimes refer to these multiple time series as "aligned time series" [Huhtala et al. 1999].

In multiple aligned time series, each time series represents a variable. Multiple aligned time series are several connected time series of $S_1, S_2, \ldots, S_m$, where $S_1 = \{s_{11}, s_{12}, \ldots, s_{1n}\}$, $S_2 = \{s_{21}, s_{22}, \ldots, s_{2n}\}$, through $S_m = \{s_{m1}, s_{m2}, \ldots, s_{mn}\}$. On the contrary, a multivariate time sequence is "a series of data elements, each element being represented by a multidimensional vector" ([Lee et al. 2000], page 599). Lee et al. [2000] treated video stream and image data as multvariate data sequences composed of several video frames, each of which has a number of attributes such as color, shape, and text. The fact that a video frame has several variables at each time point makes the video stream multivariate. Therefore, a multivariate time sequence of a video frame is a time series S where $S = \{s_1, s_2, \ldots, s_n\}$, and s_i, where $i = 1$ to n, is a feature vector of a video frame. Lee et al. [2000] applied the multivariate data sequence structure to the task of retrieving similar video sequences such as from TV news, dramas, and documentaries. Instead of a sequential search, they used minimum bounding rectangles (MBR) to represent the data structure and were able to achieve 16–28 times faster retrieval.

3. DISCRETE WAVELET TRANSFORMATION

Discrete wavelet transform possesses many favorable properties that are useful for researchers in the time series data mining field; therefore, it is essential to understand the foundation of DWT in order to appreciate its usefulness and fully comprehend the application of DWT on time series data mining. This section contains an introduction to DWT and the benefits and functionalities of DWT on time series analysis and mining.

3.1. Introduction to Discrete Wavelet Transformation

Discrete wavelet transform transforms a time series using a set of basis functions called *wavelets*. The purpose of transformation is to reduce the size of data and/or to decrease noise. By name, *wavelets* mean *small waves* [Percival and Walden 2000]. Wavelets are a set of mathematical functions used to decompose data into different components. Time series data components are separated into different frequencies at different scales by DWT. In the signal processing field, frequency is the number of repeated occurrences over a unit of time. Scale is the time interval of that time series. For example, a time series with a frequency of five event occurrences per minute represents an interval (scale) of 12 s between events. Since DWT is a data transformation technique that produces a new data representation which can be dispersed to multiple scales, the

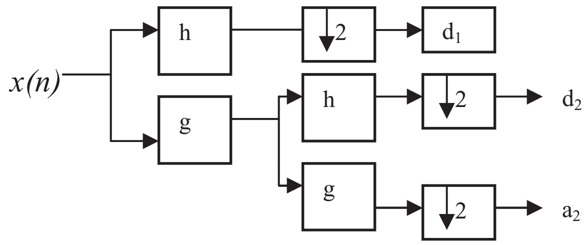

Fig. 1. A two-level decomposition.

analysis of the transformed data can be performed at multiple resolution levels as well.

Wavelet transforms analyze signals at multiple resolutions for different frequencies, as opposed to a constant resolution for all frequencies as is the case for short-time Fourier transforms (STFT). In a wavelet transform, a signal is multiplied by a wavelet function, a localized wave with finite energy, and the transform is analyzed for each segment. A continuous wavelet transform (CWT) is given by the following equation:

$$H(x) = \frac{1}{|\sqrt{\zeta}|} \int x(t) \cdot \psi^* \left(\frac{t - \tau}{\zeta} \right) dt,$$

where $H(x)$ is the wavelet transform for the signal $x(t)$ as a function of time (t), ζ is the scale parameter, τ is the time parameter, and ψ is the mother wavelet or the basis function with * denoting the complex conjugate. The scale parameter corresponds to the frequency information and equals 1/frequency and either dilates or compresses the signal. High frequencies (equivalently, small scales) compress the signal and provide global information, whereas low frequencies (large scales) dilate the signal and provide detailed information hidden in the signal. The time parameter is shifted along the signal and provides location information.

The computation of a CWT is done using wavelet series by sampling from the time-scale plane. However, it is still very expensive and DWT provide an efficient computation by using subband coding where the signal is passed through filters with different cutoff frequencies at different scales. The DWTs is computed by successively passing a signal through high-pass and a low-pass filters, producing detail and approximation coefficients. The half-band filters down-sample the signal by a factor of 2 at each level of decomposition. This generates a decomposition tree known as *Mallat's decomposition tree*, shown in Figure 1, where $x(n)$ is the signal, h and g are the high- and low-pass filters, respectively, and d_1, d_2, and a_2 are the first- and second-level detail and the second-level approximation coefficients, respectively. This approach of decomposition and filtering can be repeated until the desired level has been reached. The original signal can be reconstructed from the approximation and detail coefficients at every level by up-sampling by two, passing through high- and low-pass synthesis filters, and adding them.

A number of basis functions exist that can be used as the mother wavelet. The characteristics of the transformation are impacted by the choice of the mother wavelet, and thus the application requirements should be taken into consideration in choosing the mother wavelet. The oldest and simplest wavelet is the Haar wavelet, where the mother wavelet can be described as follows:

$$\psi(t) = \left\{ \begin{array}{rl} 1 & 0 \le t < \frac{1}{2} \\ -1 & \frac{1}{2} \le t < 1 \\ 0 & otherwise \end{array} \right\}.$$

A second group of wavelets, called *Daubechies wavelets*, offers a family of orthogonal transforms that offer a maximal number of vanishing moments for a given support. Daubechies wavelets are not expressed in terms of the resulting scaling and wavelet coefficients and cannot be expressed in closed form. Daubechies wavelets range from Daub2 to Daub22, where the index refers to the number of coefficients which doubles the number of vanishing moments. They are used in a broad range of problems such as signal discontinuities and self-similarity properties in signals. Other wavelets include Symlets, Coiflets, Meyer, Morlets, and Mexican hat wavelets. Of these Meyer, Morlets, and Mexican hats are symmetric, which possess many desirable properties for edge localization in images.

3.1.1. Calculating DWTs. For unfamiliar readers, we portray an explanation of the concept of DWT and its multiscale transformation through a simple example below: a time series S with a length of $N = 8$, consists of eight data points, each denoted with S_i, when $i = 1$ to 8, with the following values:

80 61 75 71 63 59 76 63

Then we use DWT to separate the time series S into two components (averages and differences) by calculating the pairwise averages of data points within S while preserving the pairwise differences between the data points.

The first level of transformation, which is derived by applying a Haar wavelet (the simplest wavelet function) to S, is exhibited below. The averages are presented in bold and the differences are presented in italics.

70.5 **73** **61** **69.5** *−9.5* *−2* *−2* *-6.5*

To obtain the above result, we simply apply the pyramid algorithm of Haar wavelet transform [Mallat 1989]. The first number in bold from the left is derived by adding the first two consecutive numbers from the original time series S, and then dividing the sum by 2, that is, $(S_2 + S_1)/2$. The second number in bold is derived by adding the next two consecutive numbers from S, then dividing the sum by 2, that is, $(S_4 + S_3)/2$. This averaging operation continues until the algorithm reaches the last number of the original time series. This process results in the four average numbers in bold, for a time series with a length of 8.

The first italic number from the left is derived by subtracting the first number of the original time series S from the second number, and then dividing the difference by 2, that is, $(S_2 - S_1)/2$. The second italic number is derived by subtracting the third number from the fourth number of S, then dividing the difference by 2, that is, $(S_4 - S_3)/2$. This difference operation continues until the algorithm reaches the last number of the original time series. As a result of this process, we derive four differences in italics, for a time series with a length of 8.

In general, the average numbers are derived by a shifting function, $(S_{n+1} + S_n)/2$, along the pairwise data of the original time series and the differences are derived by another shifting function, $(S_{n+1} - S_n)/2$, along the pairwise data of the original time series. The values in bold are therefore called *wavelet approximation coefficients* and the values in italics are called *wavelet detail coefficients*. More concisely, the number of wavelet approximation coefficients from the first transformation of the original time series with a length of N is $N/2$, and so is the number of wavelet detail coefficients.

Then, considering only the approximation coefficients (we leave the wavelet detail coefficients alone), we produce a second set of transformation from our original time series S by reapplying the Haar wavelet function to the four wavelet approximation

S	80	61	75	71	63	59	76	63
Level 1	**70.5**	**73**	**61**	**69.5**	*-9.5*	*-2*	*-2*	*-6.5*
Level 2	**71.75**	**65.25**	*1.25*	*4.25*				
Level 3	**68.5**	*3.25*						

S	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8
Level 1	$\mathbf{A}_{11}$	$\mathbf{A}_{12}$	$\mathbf{A}_{13}$	$\mathbf{A}_{14}$	D_{11}	D_{12}	D_{13}	D_{14}
Level 2	$\mathbf{A}_{21}$	$\mathbf{A}_{22}$	D_{21}	D_{22}				
Level 3	$\mathbf{A}_{31}$	D_{31}						

Fig. 2. The original time series data and three levels of wavelet transformed data (top) and the notation of the original time series data and the notation of wavelet transformed data (bottom).

coefficients, resulting in

71.75 **65.25** *1.25* *4.25*

Again, we have wavelet approximation coefficients in bold and wavelet detail coefficients in italics. The second set of transformation is derived only from the first set's wavelet approximation coefficients. Consequently, the number of wavelet approximation coefficients from the second transformation of the original time series with a length of N is $N/4$, and so is the number of wavelet detail coefficients. The wavelet approximation coefficients of length $N/2$ from the first transformation are decomposed into both wavelet approximation coefficients and wavelet detail coefficients of length $N/4$ each.

We can still reapply the Haar wavelet function to our second set of wavelet approximation coefficients one last time.

68.5 *−3.25*

As a result, the number of wavelet approximation coefficients from the third transformation of the original time series S with a length of N is $N/8$, and so isthe number of wavelet detail coefficients.

Note that we can easily reconstruct these approximation and detail coefficients into the original time series S. For example, S can be perfectly reconstructed given the approximation coefficients and the detail coefficients from the first transformation. At the same time, S can also be reconstructed perfectly given approximation coefficients from the second transformation, the detail coefficients from the second transformation, and the detail coefficients from the first transformation. This is because approximation and detail coefficients from the second transformation are the results of decomposing approximation coefficients from the first transformation.

Denoting the first, second, and third transformations as level 1, 2, and 3, respectively, we can specifically claim that the second-level approximation coefficients, **71.75** and **65.25**, can be reconstructed without a loss of information from the third-level approximation and detail coefficients, **68.5** and *−3.25*, by applying an inverse Haar wavelet transformation. Therefore, given (1) the approximation coefficient from the last level, **68.5**, (2) the detail coefficients from every level, and (3) the wavelet function used in the transformation, one can easily reconstruct the original time series S. Figure 2 summarizes the original time series data and its transformations from the example.

The approximation coefficient from the last level of transformation contains the most important information of time series as it summarizes the time series. Thus researchers can choose to discard other less important components if they need to. For

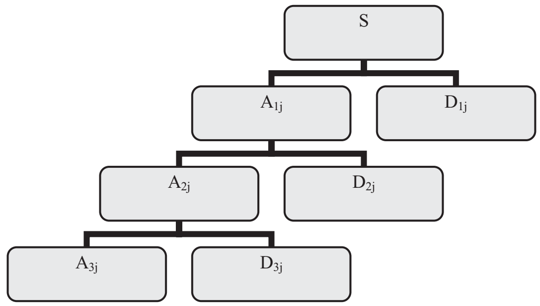

Fig. 3. Multi-level decomposition tree by wavelet transforms.

this reason, DWT is utilized for data reduction of time series in order to save storage space while sacrificing a small amount loss of detail information. By and large, the last approximation coefficient and few of the high-level detail coefficients are usually selected for preservation. DWT is therefore regarded by researchers as a lossless transformation, whereby data from the transformed domain can collectively reconstruct the original time series data. The only coefficients needed for a perfect reconstruction are the approximation coefficients from the last level of transformation and the detail coefficients from every level of transformation.

Figure 3 depicts the transformation of the original time series S and its wavelet transformation. A_{ij} denotes wavelet approximation coefficients and D_{ij} denotes wavelet detail coefficients, where i denotes the level of transformation and j denotes the order of wavelet coefficients. DWT decomposes a single signal into multiscale signals using wavelet functions. Consequently, DWT is considered as a time-scale transformation [Misiti et al. 2005]. Each decomposed signal component is still in the time domain, rather than in other domains, and is dispersed into different scales.

3.1.2. Benefits of DWTs. DWT is a very useful technique for time series data processing of many aspects such as data dimensionality reduction, noise reduction, and multiresolution analysis. In the signal processing field, one can take an original signal and distribute it into separate signals in different frequencies by applying a wavelet function, while preserving the original signal, which can be reconstructed from these separate signals. For time series data, DWT can create separate time series from the original time series. The original information will be distributed into these different time series in the form of wavelet coefficients. Therefore, DWT is considered an orthonormal transformation, meaning it allows reconstruction and preserves the original information (also known as *energy*) of the original signal within the transformed data. The function of DWT as an orthonormal transformation is to reduce the high dimensionality of a time series into a much more compact data representation, with complete information stored within its coefficients. Therefore, DWT is suitable for analyzing time series data for the following reasons.

First, DWT is an effective method for time series data reduction. As mentioned before, each individual time series is composed of continuous observations of an interesting phenomenon, and therefore is likely to be very large. Fortunately, DWT lends itself very well to time series data analysis because it is very effective in reducing large time series data into a significantly smaller number of coefficients, as confirmed by many studies [Chan and Fu 1999; Liabotis et al. 2006; Popivanov and Miller 2002; Wu et al. 2000]. Researchers can utilize DWT to project a large time series into DWT coefficients, and then perform other data analyses on these coefficients. For similarity search applications, performing analyses on the coefficients will likely generate more false retrieval than performing analyses on the original time series. Yet, if the methods follow the lower-bounding condition of the GEMINI framework, false hits can be pruned in the postprocessing. The efficiency gains obtained from a reduced dimensionality are worth the effort of pruning more false hits.

Second, DWT can detect sudden signal changes well, because it transforms an original time series data into two types of wavelet coefficients: approximation and detail. Approximation wavelet coefficients capture rough features that estimate the original data, while detail wavelet coefficients capture detail features that describe frequent movements of the data. Researchers can investigate the latter and discover sudden changes, peaks, or spikes in the observed phenomena. These sudden changes are sometimes difficult to detect in the original data because they are obscured by an overall trend or seasonal movements of the data. Moreover, detection can be a time-consuming task if performed solely by human experts. DWT can help relieve the burden of detection by separating detail features from the original time series, and thereby sudden changes or spikes can be uncovered easily. Subsequently, researchers are free to apply various detection techniques to the data that is proposed in the literature.

Third, DWT is useful in supporting multiresolution analysis. In addition to projecting a time series into approximation and detail wavelet coefficients, DWT decomposes these coefficients into various scales. When $S1$ and $S2$ are two resultant time series with different time scales and b is a scaling factor between them, $S1_i = S2_{(i*b)}$, for $1 \leq i \leq N$, where $S1_i$ and $S2_i$ denote the locations of event in time. A scale reflects a time interval within a time series. This allows researchers to analyze wavelet coefficients from one temporal scale individually, as well as to choose multiple temporal scales to be investigated collectively.

Altogether, DWT is extremely powerful for data reduction and signal compression because of its orthonormal property. The application of DWT has been studied in several areas, such as image compression, noise filtering, and singularity detection. Its benefits warrant a further investigation to search for domain applications where the properties of DWT suit their purposes. In conclusion, DWT possesses many capabilities that have a large potential for supporting novel data mining approaches for time series data.

We will briefly discuss the functionalities of DWT for data mining in a variety of application domains. In brief, they are dimensionality reduction, noise filtering, and singularity detection.

3.2. Discrete Wavelet Transform for Dimensionality Reduction

One of the main reasons for data transformation is data reduction, which is a crucial preprocessing step of data analysis and mining. A data reduction step before applying data analysis and mining enables faster execution of the algorithms, since it reduces the size of the original time series, therefore lowering the access time to data.

In order to reduce time series dimensionality using DWT, only some wavelet coefficients are retained in data mining systems. This calls for an important decision of which coefficients to drop. One of the most common and popular ways to approach this problem is to retain a few coefficients which contain the most energy and drop the remaining ones. Wavelets' energy is a statistics calculated from wavelet coefficients. One may obtain this information by plotting the distribution of values among coefficients. Once these coefficients with high energy are identified, other coefficients can be dropped. An excellent example of a coefficient-dropping strategy can be found in Shahabi et al. [2000]. Shahabi et al. [2000] not only employed the above approach, but also proposed other approaches to strategically drop wavelet coefficients for their OTSA tree. They first proposed dropping the nodes which contain less energy to reduce a disk space requirement of the OTSA tree. However, dropping nodes means losing potential outlier information. Since their work was originally intended for multilevel trend and surprise queries, they suggested keeping those outliers in a more condensed space in a form of position-value pairs instead of within tree nodes. If the system's space is still limited, some coefficients will need to be abandoned and those coefficients' energy will be lost. In that case, there are two decisions which can be made. First, the head coefficients can be retained and tail coefficients dropped. Second, the high-energy coefficients can be retained and low-energy coefficients dropped. The OTSA tree can make these decisions on-the-fly depending on the available disk space.

It is possible to retain only a few coefficients for the similarity search task for dimensionality reduction, although one still needs to guarantee no "false dismissals." It needs to be verified that data reduction techniques are lower-bounded in order to comply with the GEMINI framework. Orthonormal transformations always follow a lower-bounding property, also known as a *contractive property* [Keogh et al. 2001]. Therefore, any dimensionality reduction technique that is an orthonormal transformation automatically conforms to the GEMINI framework and will be able to further reduce the similarity search time, while preserving accuracy. These dimensionality reduction techniques transform time series data into another format. The information is compressed into Fourier coefficients for discrete Fourier transform (DFT) and into wavelet coefficients for DWT, where most of the information is squeezed into a few coefficients. Empirically, these techniques have helped reduce a long time series of an original dimensionality of 1024 to a transformed dimensionality of 16–20 [Liabotis et al. 2006]. This is a data reduction of two orders of magnitude. In addition, there is no false dismissal from a similarity search task of an orthonormal transform of DFT, SVD, or DWT. Following the GEMINI framework, similarity search on the transformed data is likely to result in retrieving some false hits. However, false hits will be pruned in the postprocessing step of a similarity search task. Nevertheless, the time saved by dimensionality reduction outweighs the pruning time of false hits.

Since DWT was proposed as a dimensionality reduction technique for the similarity search task by Chan and Fu in 1994, there have appeared several followup studies on utilizing DWT for similarity search [Chan and Fu 1999; Popivanov and Miller 2002]. A comparison between DFT and DWT was reported in the literature by Wu et al. [2000], who found that DFT and DWT produced a marginal difference as dimensionality reduction techniques for a similarity search task. The query-matching error was not significantly improved, and DWT did not increase query-matching precision. However, another study by Liabotis et al. [2006], as well as the study of Chan and Fu [1999], reported that DWT outperformed DFT. These empirical evaluations imply that DWT can perform equally or better than DFT as a dimensionality reduction technique for similarity search.

Another research area closely related to dimensionality reduction in similarity search is the application of DWT in several fields of data compression. Examples include studies in signal compression and image compression [Castelli and Kontoyiannis 1996, 1999; Castelli et al. 1996]. The applications of dimensionality reduction in time series also include two-dimensional image classification, where images are compressed but still retain enough information to be distinguished among classes [Brambilla et al. 1999; Chang and Kuo 1993; Jacobs et al. 1995].

3.3. Discrete Wavelet Transform for Noise Filtering

Another useful function of DWT is noise filtering. Noise is usually identified by domain experts to be high variations of data mixed into real signals [Han and Kamber 2006; Orfanidis 1996]. The basic idea of noise filtering is to isolate noise (unwanted signals) from true information (wanted signals). Therefore, a suitable technique for noise filtering must have the ability to separate and isolate the noise from the signal.

DWT is a suitable technique to filter out noise because, when a signal or a time series is decomposed by DWT, the original signal is separated into approximation and detail coefficients at different resolution levels. The information of the original signal is retained in wavelet coefficients and a perfect reconstruction of the original data can be performed from these coefficients. However, some of the detail coefficients, which represent detail movements in data, may be recognizable as noise. Those coefficients can then be set to zero prior to a DWT reconstruction process in order to filter out noise from the original time series. In other words, the reconstruction involves rebuilding a time series from every component but noise.

This DWT functionality enhances the capability of various data analysis and mining applications. Classification can be performed on a noise-filtered signal better than on a noise-blended signal. For example, better classification results were reported by Subasi [2005] in classifying epileptic and normal patients when applying DWT. Dinh et al. [2002] also reported that wavelet-based features can be dependably employed in audio genre classification for better classification results.

3.4. Discrete Wavelet Transform for Singularity Detection

Singularity is normally a point where the time series signal behaves irregularly. Such behavior usually reveals interesting information in the time series signals [Mallat and Hwang 1992]. For example, in images, adjacent pixels with extremely different densities inform us of a picture edge. The location of these edges is useful for image recognition.

Singularity detection involves the analysis of transient events in the form of peaks or cusps [Mallat and Hwang 1992]. Since DWT decomposes time series data into elementary components, it is straightforward to detect local regular and irregular structures. Using DWT, it takes minimal effort to detect any bursts, cusps, or irregularities in data. A study that utilized DWT's time aspect in the detection of jumps and sharp cusps in time series appeared in Wang [1995]. A close examination of each scale helped detect any spikes, which otherwise might be left unnoticed in the original signal. In general, spikes are considered as pointed-end parts of the signal. These pointed-end sections of signals can also be called *cusps*. From this study, spikes or cusps, which present quick local variations in signals, were shown to be enhanced by DWT through wavelet detail coefficients [Struzik and Siebes 2000; Wang 1995].

As previously described, DWT is a time-scale transformation because each scale contains transformed data in time domain, rather than in other domains. The components at each different scale are separated by their periodicity. One can detect short-time phenomena in one or more scales of the wavelet-transformed data, whose multiscale detail coefficients are an indicator for multilevel surprises [Shahabi et al. 2000]. Such an ability to detect abrupt changes is valid for both one-dimensional and two-dimensional time series data. Therefore, DWT becomes a popular and powerful technique for the image processing field as well. For example, the ability of DWT to detect edges in images and textures has been exploited for image recognition and progressive image classification. We will discuss more about the studies that apply DWT's singularity detection to the domain applications in Section 4.3.

In addition, wavelet coefficients have a time-localization property, which will be explained below [Dillard and Shmueli 2004; Percival and Walden 2000]. Other frequency-related techniques, such as discrete Fourier transform (DFT), present information about the events of interest in the frequency domain but lose information in the time domain. Unlike frequency-related transforms, DWT preserves temporal information—the information regarding when the events occur. Let us examine an example recreated from the Wavelet Toolbox User's Guide [Misiti et al. 2005] which illustrates an apparent difference between DFT and DWT in singularity detection.

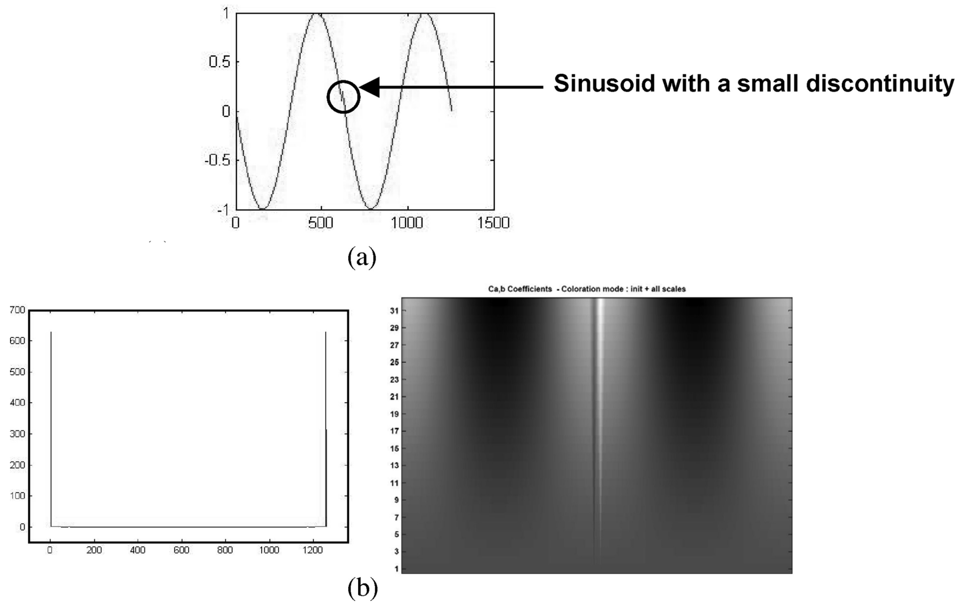

Fig. 4. (a) A synthetic signal example [Misiti et al. 2005]. (b) The analyses of the synthetic signal using DFT (left) versus DWT (right) [Misiti et al. 2005].

In this example, a synthetic time series with a single small discontinuity is created. The discontinuity is extremely tiny, so much that it is invisible with the bare eyes at this scale. When this signal is transformed using DFT, the x axis denotes amplitude and the y axis denotes frequency of the signal. The DFT graph is a flat spectrum with two peaks at the ends of the x axis which represents a mostly single frequency with high frequencies at two amplitudes [Misiti et al. 2005; Shasha and Zhu 2004]. In Figure 4(b) left, the DFT plot does not tell us the existence of a discontinuity in the synthetic signal. However, DWT coefficients are plotted with the x axis denoting time and the y axis denoting scale. The color at each pixel illustrates the intensity of wavelet coefficients at each time snapshot of a specific scale. In Figure 4(b) right, the exact location of the discontinuity is clearly shown at the bottom of the plot. Not only does DWT show the existence of the discontinuity by the distinctly different intensity from its local neighborhood, but it also points out the discontinuity (Figure 4(b), right) at the same time location as the synthetic data plot (Figure 4(a)). In brief, DWT is a transformation technique with a time-localization property.

In general, detection of spikes or cusps using DWT can be exemplified by plotting data and inspecting for any visual clues. Examples of applications that have employed the singularity detection by DWT include anomaly detection in credit card transactions, underwater signal detection [Bailey et al. 1998], anomaly detection in computer networks [Huang et al. 2001; Magnaghi et al. 2004], intrusion detection [Lee and Stolfo 1998], disease outbreak detection [Shmueli 2004; Wong 2004], and anomaly detection in physiological monitoring data [Saeed and Mark 2001]. Consequently, DWT can help researchers in detecting anomalies for various application domains.

4. WAVELET-BASED TIME SERIES DATA ANALYSIS AND MINING

In Section 3, we introduced DWT, its properties, and its favorable characteristics. DWT has gained an increasing popularity in the past few decades and there is a significant amount of work in this field. DWT is an effective and powerful instrument for a time-scale analysis of nonstationary signals that can be found in the medical and biological domains, computer network traffic data [Basu et al. 1996; Huang et al. 2001; Kobayashi and Torioka 1994; Ma and Ji 1999a, 1999b; Ma and Ji 2001; Magnaghi et al. 2004; Riedi et al. 1999], multimedia archives, and the chemical engineering field [Chen et al. 1999a, 1999b]. For example, signals in the medical and biological domains include EEG (electroencephalograph signal) [Subasi 2005], myocardial tissue image [Mojsilovic et al. 1997], heartbeat and carbon dioxide level [Nilsson et al. 2005], and other clinical data [Goodwin and Maher 2000; Rizzi and Sartoni 1994; Silver and Ginsburg 1984]. Multimedia archives include digital images [Ardizzoni et al. 1999; Blume and Ballard 1997b; Brambilla et al. 1999; Chang and Kuo 1993; Jacobs et al. 1995; Mandal et al. 1999; Mojsilovic et al. 1997; Natsev et al. 1999; Sheikholeslami et al. 1999; Wang et al. 1997a, 1997b, 1997c], audio signals [Dinh et al. 2002; Lambrou et al. 1998; Li and Khokhar 2000; Subramanya and Youssef 1998; Tzanetakis and Cook 2002; Tzanetakis et al. 2001], satellite images [Castelli et al. 1996], and video streams. A number of researchers have realized and utilized this multiresolution property of DWT to enhance the data analysis and mining process in their research, as will be discussed next.

The multiresolution analysis presented by DWT has a unique and important characteristic. The decomposition of an original signal into different scales separates hidden but meaningful subsignals of the original time series. The decomposed signals, which convey the information more meaningfully, are then fed into wavelet-based data analysis and mining techniques, such as in the medical domain [Subasi 2005] and in network traffic data [Huang et al. 2001].

To fully appreciate the benefits of DWT for time series data analysis and mining, we now focus on using DWT for time series data analysis and mining tasks. Briefly, the tasks can be separated into five categories: similarity search, classification, clustering, detection, and prediction.

4.1. Wavelet-Based Similarity Search in Time Series

Many applications with temporal data require the capability of searching for similar or exact time series. Since the late 1980s, time series research has shifted from focusing on matching exact time series to searching for similar time series [Agrawal et al. 1993], the latter being referred to as a *similarity search*. Similarity search of time series needs a parameter called a *distance tolerance*, which is classified as follows.

(1) *Exact matching*. Distance tolerance = 0
(2) *Similarity search*. Distance tolerance = a threshold ε, which is set by users.

The distance tolerance between two time series signifies the upper bound of the distance between the two time series. An exact matching is a special case of time series similarity search, where the distance tolerance between two time series is zero. Therefore, a retrieved time series resulting from an exact matching operation has to be zero distance units apart. Similarly, a retrieved time series resulting from a similarity search operation will be less than or equal to ε distance units apart from the query sequence. The search, either exact or similar, is applied in situations such as identifying music scores with the exact same sequence of notes, and identifying stocks with similar price growth.

Time series search can also be classified by the length of time sequence queried (query sequence) compared to time sequences already existing in the database (archive sequence) [Agrawal et al. 1993; Faloutsos et al. 1994; Keogh et al. 2001] as follows.

(1) *Whole matching*. The length is equal to the length of archive sequences.
(2) *Subsequence matching*. The length is shorter than the length of archive sequences.

To perform a similarity search task, long time series are partitioned into smaller sequences and then stored in the database. A new unknown sequence called a *query sequence* is to be compared with existing archive sequences in the database in order to find either an exact or the most similar time sequence. The sequences to be compared in whole matching have the same length, while those in subsequence matching have a smaller length for the query sequence. The shorter length sequence is matched against the longer length sequence. As an example, whole matching may help find copyrighted music in a song archive. Subsequent matching facilitates meteorologists in recognizing critical weather patterns that need immediate action.

For a similarity search task, there are two main steps, indexing and query processing. *Indexing* is the process of constructing pointers for faster access to data. An index is computed from values within the time series. Since time series data has a very high dimensionality, direct indexing is computationally expensive and becomes unaffordable [Popivanov and Miller 2002]. A feature extraction needs to be done prior to index construction, in order to reduce the dimensionality of an index. *Query processing* is the process of finding the matches of time series. This process involves developing or utilizing similarity measurements, which quantify how alike two time series are. Query processing compares a collection of time series using selected similarity measurements, then retrieves the time series from databases using an index, and returns the query results to the users.

There are two aspects to consider for the performance of a similarity search: speed and precision. Speed indicates how fast the similarity search completes its task. The overall time taken for a similarity search task is comprised of (a) a time for constructing and updating an index, (b) a time for searching for similar time sequences, and (c) a post processing time. The latter two are under the query processing step. Searching for similar sequences is likely to result in retrieving some false hits at first. The postprocessing step prunes false hits and results in more accurate results. Precision measures the accuracy of the retrieved time series. It equals the number of correctly retrieved time series, divided by the number of total retrieved time series. In other words, it is the percentage of the retrieved time series that are similar.

The most straightforward measurement for evaluating speed is time. However, differences in hardware and system configurations may lead to different reporting times, and therefore, the most reliable measurement for evaluating the speed performance of similarity search techniques is data page accesses [Popivanov and Miller 2002], which measures the number of disk page reads. Data page accesses are further broken down into index page accesses and time series data page accesses.

The research interest in the area of time series similarity search mostly focuses on feature extraction, indexing, and similarity measurements. Indexing methods specially designed for multidimensional data have been proposed and explored for similarity search. These indexing methods include the quadtree, the grid structure and grid file, the R-tree family, and the KD-Btree family [Agrawal et al. 1993; Shasha and Zhu 2004]. Such indexing methods support high dimensionality of time series data by extending binary tree index to higher dimensions in different fashions. As for grid structure and grid the file, the index memory is divided into a d-dimensional space and each d-dimensional data point is hashed into each individual grid cell.

Even though all of the above indexing methods are especially designed for high-dimensional data, empirical results have shown that the number of dimensions still largely affects the query time [Agrawal et al. 1993]. Most of the proposed indexing approaches scale exponentially in terms of computation time for high dimensionalities. Due to the aforementioned "dimensionality curse" problem, there is a limit to the number of dimensions that allows indexing methods to work well. Spatial and high-dimensional indexing methods perform efficiently up to 8 to 12 dimensions, while most time series queries range from 20 to 1000 time points [Keogh et al. 2001; Shasha and Zhu 2004]. To get around this problem, researchers have called for feature extraction or dimensionality reduction techniques in order to reduce the dimensionality to a manageable number for indexing.

Feature extraction techniques have also been explored in the literature on time series [Man and Wong 2001; Yoon et al. 2005]. Feature extraction aims to transform data into a lower dimensionality by removing redundant information while preserving the time series' most important elements [Chen et al. 1999a, 1999b]. Dimensionality reduction also focuses on trimming down the number of dimensions of time series [Chakrabarti et al. 2002; Keogh et al. 2001]. Both dimensionality reduction and feature extraction serve as a subtask of a similarity search task: during a similarity search, researchers choose an appropriate feature extraction or dimensionality reduction technique to ensure that the extracted features have sufficient information for similarity search algorithms to differentiate between time series. As a result of these methods, the number of dimensions is reduced, keeping only a few numbers for indexing and query searching. However, researchers must make a difficult choice about the optimal number of dimensions to be retained. There is a tradeoff between speed and precision performance.

This speed and precision tradeoff has always framed the research direction for similarity search indexing aiming at similar time series not only regarding correctness but also regarding acceptable computation time. In addition, it also has to meet the accuracy constraint.

Chan and Fu [1999] proposed the use of DWT for reducing the dimensionality of time series. The potential of DWT for this purpose had been pointed out previously by Korn et al. [1997]. According to Chan and Fu [1999], DWT representations of a time series carry information about both time and frequency locations, whereas DFT representations fail to encapsulate the former information. Chan and Fu [1999] mathematically proved that DWT produced data representations that obeyed the lower-bounding condition of the GEMINI framework, which guarantees that no eligible similar time series is discarded. Experiments conducted for this study showed results that DWT considerably outperformed DFT in terms of precision and the number of page accesses. Moreover, the scalability of DWT was shown to be better than that of DFT when increasing the database size and the time sequence length.

According to Li et al. [2003], there are three approaches to applying DWT to the similarity search task: keeping the first few wavelet coefficients, extracting features and defining new similarity measures using wavelets, and supporting similarity search in a multiscale fashion.

In the first approach, Chan and Fu [1999] proposed DWT to map an n-dimensional time series into a k-dimensional space for similarity search. In this study, the first few wavelet coefficients were kept to retain most of the information. In addition, Wu et al. [2000] took the above DWT-based approach and compared it with a DFT-based approach. The result from Wu et al. [2000] did not show a significant difference between these two approaches for similarity search in terms of matching precision, but DWT naturally has significantly lower time complexity than DFT [Li et al. 2003].

In the second approach, Struzik and Siebes [1999a, 1999b] proposed the use of a special data representation that preserved only the sign of wavelet coefficients, instead of the first few coefficients as in other studies [1999a, 1999b]. The sign information gives relative values of wavelet coefficients, which is more useful in the case of comparing similarity among time series than absolute values. Beyond the studies of Struzik and Siebes, other studies have devised DWT to extract compact feature vectors and defined new similarity measures and new indexing schemes to accommodate the search [Jacobs et al. 1995; Li and Khokhar 2000; Wang et al. 1997a]. Jacobs et al. [1995]

distilled wavelet coefficients into small signatures for each image. Also, a new similarity measure called the *image querying metric* was developed to compare common significant coefficients between the query images and the target images. The image querying metric was proven effective both in terms of speed and success rate in querying a large image database in Jacob et al.'s [1995] article. Wang et al. [1997a] utilized a combination of wavelet features, which included Daubechies' wavelets, normalized central moments, and color histograms, to create a new vector for image similarity matching. The new developed feature vector allowed searching by partial sketch images for large image databases. Li and Khokhar [2000] exercised the knowledge that the wavelet decomposition of audio sounds is extremely like the decomposition in sound octaves. By including several statistical properties of wavelet decomposition, such as zero crossing rate, mean, and standard deviation, in their hierarchical indexing scheme, they reported a great recall rate of more than 70% on a set of diverse audio signals. In sum, the studies within this second approach of wavelet-based similarity search were feasible and practical, owing to the researchers' use of heuristics in devising new meaningful features and associated similarity measurements from their domain knowledge.

In the third approach, DWT is used in finding similarity in a step-by-step manner for several data formats, such as images [Natsev et al. 1999] and audio sounds [Li and Khokhar 2000]. Brambilla et al. [1999] exploited a DWT's favorite functionality, multiresolution analysis to describe images. By applying DWT, four subimage wavelet coefficients—one approximation and three details—are produced as a result of the original image. Then a wavelet decomposition is reapplied to the approximate image of the first level to obtain the next four subimages at the next level. This process is repeated on the approximate image at each level. By keeping the 128 largest wavelet coefficients and setting the rest of the coefficients to zero, an image signature is formed. This image signature describes pictorial content and captured sufficient perceptual similarity between images, which the human eye would use in image recognition. Another example of image similarity search in a multiresolution fashion was a study by Jacobs et al. [1995]. Multiresolution wavelet coefficients of an image provide independent information at each level to the original image. This information is distinct in terms of color shift, poor resolution, dithering effects, and misregistration. Therefore, this multiresolution image retrieval method allows for querying the target image using these distinctive image features extracted from multiresolution wavelet coefficients. Also, Struzik and Siebes [1999a, 1999b], as discussed earlier under the second approach, utilized a multiscale representation of time series as well. Multiscale wavelet coefficients permit the construction of a scale-wise hierarchical organization of wavelet extracted features. Such a hierarchical structure facilitates a stepwise comparison on time series via their correlations. The comparison is based on this special wavelet representation, which is otherwise unavailable if DWT is not employed.

In conclusion, the multiresolution property of DWT is inherent and offers an opportunity for a similarity search to be performed on time series data at different levels of resolution. The similarity search can be performed directly on few selected wavelet coefficients, or on extracted new features. Also, the similarity search can be applied in a stepwise manner for multiscale analysis.

4.2. Wavelet-Based Time Series Classification

The goal of time series classification is to assign a class label to a time series from two or more predefined classes. Application domains for time series classification are varied. Examples of application domains include speech recognition, gesture recognition, intrusion detection [Zeira et al. 2004], audio classification [Lambrou et al. 1998; Tzanetakis and Cook 2002; Tzanetakis et al. 2001], image classification [Blume and Ballard 1997b; Castelli et al. 1996; Wang et al. 1997a, 1997b], texture classification

[Chang and Kuo 1993; Laine and Fan 1993; Scheunders et al. 1998], and medical signal classification [Subasi 2005].

Researchers consider several evaluation criteria to assess classification approaches. Accuracy is likely the most important criterion in the classification literature since the main goal is to correctly classify an unknown instance of time series data. Regarding the accuracy evaluation, researchers may also use an error rate to measure accuracy in a complementary way.

When comparing several competing classification approaches, accuracy is usually evaluated together with other criteria. The computation time (speed) of the classification algorithm is probably the second most important criterion, especially for time series data. Speed is usually reported via experimentation by measuring the clock time it takes to complete a classification task, along with reporting the computational complexity of the algorithm.

DWT provides an effective way to isolate nonstationary signals into signals at various scales. We sometimes call these signals *signal decompositions*. Various aspects of nonstationary signals such as trends, discontinuities, and repeated patterns are clearly revealed in the signal decompositions. Other signal-processing techniques are not as effective in isolating all of these transient features present in nonstationary signals. For those reasons, DWT is a suitable technique to combine with classification approaches in order to categorize an unknown signal into a predefined group of signals. This section explains how DWT assists in the classification process.

DWT can be integrated into a classification of time series data in two main ways. First, the classification methods are applied to the wavelet domain of the original data. Second, the multiresolution property is incorporated into the classification procedures to facilitate the process [Li et al. 2003].

The first approach is straightforward. It is simply performing a classification on wavelet-transformed data instead of the original data. A potential research question is which levels of signal decomposition to choose from, since the application of DWT produces a number of signal decompositions. The answer to this question can vary depending on the data and application domain. The second approach is more complex. DWT is utilized in classification in a progressive fashion, which means it gradually categorizes time series data from decompositions of lower resolution to higher resolution. Progressive classification serves the purpose of faster computation when the classification is performed on a much smaller set of wavelet coefficients from a coarser level of resolution. If executed in a distributed environment, progressive classification can provide a faster data transfer rate between terminal machines when contents are being classified progressively. We will discuss both approaches of wavelet-based classification in further detail by examining the research studies that fall under each approach.

4.2.1. Classification on a Wavelet-Transformed Domain. The first approach is applying the classification methods on the wavelet-transformed domain of the original data. Several researchers have employed this approach [Blume and Ballard 1997b; Dinh et al. 2002; Laine and Fan 1993; Lambrou *et al.* 1998; Mojsilovic et al. 1997; Scheunders et al. 1998; Subasi 2005; Tzanetakis and Cook 2002; Tzanetakis et al. 2001]. We will discuss these studies along their domain of applications: medical, texture, and audio.

In the medical signal classification domain, DWT helps researchers isolate relevant features from the original signal. In Subasi [2005], EEG signals were decomposed into subbands of different frequencies using DWT. Then wavelet coefficients were fed to a neural network for classification. Subasi's [2005] aim was to classify whether the EEG signal is considered epileptic or normal. If classified correctly, epileptic seizures in patients would be detected. Without an appropriate analysis, a seizure might remain unnoticed owing to its hidden presentation or might be confused with a stroke. After

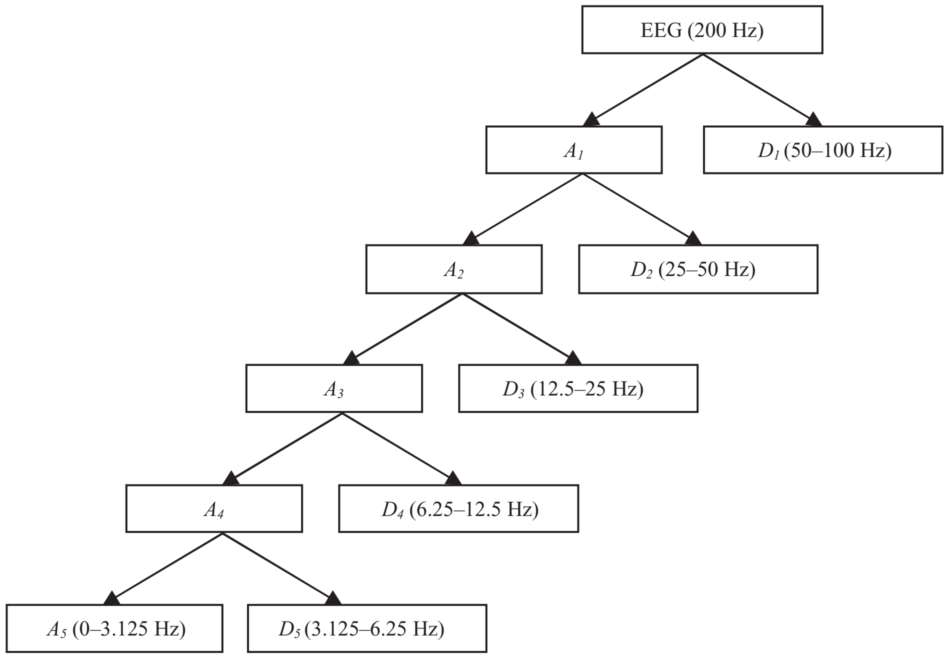

Fig. 5. Frequency decomposition of EEG signals.

DWT, only parts of the EEG signals that reside in significant frequencies are retained. The raw EEG signals are digitized at 200 samples/s (200 Hz). With the application of DWT to each signal, the corresponding decomposed signals are generated, as illustrated below. In Figure 5, A_i designates the approximation at level i of the signal, while D_i designates the detail at level i of the signal.

Different frequency ranges—δ (1–4 Hz), θ (4–8 Hz), α(8–13 Hz), and β (13–30 Hz)—convey a meaningful message to the medical experts. The resultant wavelet coefficients are also related to these meaningful subbands of the signal. For example, the components A_5 corresponds to δ (1–4 Hz), D_5 corresponds to θ (4–8 Hz), D_4 corresponds to α (8–13 Hz), and D_3 corresponds to β (13–30 Hz). To medical domain experts, the levels of decompositions that are lower than D_3 (D_2 and D_1) are not significant for classifying epileptic seizures and are therefore not included in the neural network classification model. Thus only the wavelet decompositions with corresponding frequencies to δ, θ, α, and β are extracted and fed as inputs into the artificial neural networks.

The classification accuracy of a wavelet-based neural network (DWN), an approach in also Subasi [2005], has been found to be higher than that of its counterpart technique, a regular feedforward error backpropagation neural network (FEBANN). It was elucidated by the author that, since EEG signals are nicely decomposed into the meaningful subbands of different frequencies, those extracted subbands in DWN account for better classification results, compared to the results from FEBANN.

There are a number of studies that pursued this first approach by applying the classification on the wavelet domain of the data [Blume and Ballard 1997b; Dinh et al. 2002; Laine and Fan 1993; Lambrou et al. 1998; Mojsilovic et al. 1997; Scheunders et al. 1998; Tzanetakis and Cook 2002; Tzanetakis et al. 2001; Wang et al. 1997a]. At first glance, the approach may seem straightforward, but various issues need to be taken into consideration. Theoretically, the maximum number of wavelet decomposition levels for a time series with a length of N is $log_2(N)$, and the total number of coefficients will approximately be N, that is $N-1$ detail coefficients and 1 approximation coefficient. To perfectly reconstruct the original time series, there are N wavelet coefficients to be retained, which are dispersed into $log_2(N)+1$ groups, that is, all $log_2(N)$ levels of detail coefficients and the one last level of approximation coefficient. Researchers should exclude unimportant coefficients and retain only significant components for their classification application. With such a large number of coefficients to choose from, it is almost a daunting task to make a decision about setting up parameters for classification. Most often, researchers choose the appropriate decomposition levels and the choice of wavelet coefficients with some theoretical background in the research field. We describe the following studies within different domains to perceive how decomposition levels are selected.

In the image and texture classification domain, wavelet transform revealed information about the local brightness, color, and surrounding texture features [Blume and Ballard 1997b; Scheunders et al. 1998]. Various choices of extracted features from the wavelet transformation process were applied to texture classification, as discussed below.

Blume and Ballard [1997a] interpolated between neighboring wavelet coefficients to obtain the texture information per pixel. The important feature for classification in Blume and Ballard's [1997a] study was the texture information per pixel, which was then used to obtain as high as 99% accuracy in pixel classification. In other cases, the appropriate feature to use is wavelet energy instead of wavelet coefficients [Laine and Fan 1993; Mojsilovic et al. 1997; Scheunders et al. 1998]. Wavelet energy features reflect the distribution of energy along various wavelet decomposition scales. For example, Scheunders et al. [1998] investigated texture properties such as roughness, granularity, and regularity using wavelet energy features in a multiscale manner. The multiscale analysis was applied in Scheunders et al.'s [1998] study because past studies have indicated that the human visual system processes images in a multiscale fashion. As another example, Mojsilovic et al. [1997] decomposed texture samples into each level of decomposition, then computed the energy difference between neighboring decomposition levels. The difference was then compared with a threshold so it was assured that the decomposition had not caused image degradation. Mojsilovic et al. [1997] found that the developed feature was effective in classifying clinical data, compared to other transform-based techniques.

A combination of features was also promising for image and texture classification. Recall that, for image data, the information at each pixel, that is, the location, has various features such as color and texture. When these pixels are ordered appropriately, they form a long data sequence. Sequences of image and texture information were treated as if they were time series. For instance, Laine and Fan [1993] studied texture characterization at multiple scales using both energy and entropy features. In Laine and Fan's [1993] study, the wavelet-based feature was able to classify 25 textures in the study without any error. In Wang et al. [1997b], a new feature vector for an image classification was extracted by merging Daubechies' wavelets, normalized central moments, and color histograms together. An algorithm called *WIPE*™ (Wavelet Image Pornography Elimination) was devised from this new combined feature. WIPE™ was able to classify an image as objectionable or benign. This application was particularly useful in helping the software industry to counteract the threat of pornography images.

In the audio classification domain, wavelet transformation separated audio signals into meaningful subbands of music surface and rhythm information [Tzanetakis and Cook 2002; Tzanetakis et al. 2001]. Dinh et al. [2002] decomposed audio signals at different scales using Daubechies wavelet transform. Subband signals at varying levels were characterized to meaningful different sound types. For instance, the subband signal from the first level of decomposition, which ranged from 11025 to 22050 Hz, matched with the noise and friction sound. The subband signals from the second and third levels of decomposition, which ranged from 5513 to 11025 and 2756 to 5123, respectively, corresponded with the speech and music lyrics.

Several types of features were used in wavelet-based audio classification. For one, statistics of wavelet coefficients were calculated and supplied to classifiers as in a number of papers [Dinh et al. 2002; Lambrou et al. 1998; Li and Khokhar 2000; Subramanya and Youssef 1998]. For instance, Dinh et al. [2002] proposed that the feature vector for each decomposition level of wavelet transformation is composed of the following coefficient statistics: wavelet energy, coefficient variance, zero crossing rate, centroid, and bandwidth. All of these subband features were a product of an additional computation from wavelet coefficients, and were found to be able to distinguish successfully six video genres in the study [Dinh et al. 2002]. Lambrou et al. [1998] also explored wavelet coefficient statistics, but they utilized a more extensive set of calculated statistics for their classification of audio sounds. A total of eight statistical features were collected as inputs and fed into four classifiers for comparison. Lambrou et al.'s [1998] statistics also proved superior with an empirical classification accuracy of 91.67%.

Besides audio classification, researchers extracted musical features using coefficient statistical features and their domain knowledge. Two studies by Tzanetakis, Cook, and Essl investigated music instead of audio sounds, and performed a music genre classification [Tzanetakis and Cook 2002; Tzanetakis et al. 2001]. The articles mentioned that the statistical features are related to instrumentation, rhythmic structure, and the form of genre members. These features could define a particular music genre. According to Tzanetakis et al. [2001], the characteristics of music—referred to as *musical surface*—were related to texture, timbre, and instrumentation. In this study, nine musical surface features were calculated based on the Fourier transform. In addition, musical rhythmic structure was another musical feature which was calculated using wavelet transform. Altogether, these features from Fourier and wavelet transforms were fed into classifiers to successfully define the rhythmic structure and strength of music.

4.2.2. Progressive Classification by Wavelets. The second approach to wavelet-based time series classification is to facilitate the classification process by incorporating DWT's multiresolution property. A multiresolution analysis is used to classify data—usually images or texture features—progressively. For example, Castelli et al. [1996] applied generic classifiers on a low resolution representation of wavelet-transformed data. They defined the satellite image classification as "the process of labeling individual pixels or larger areas of the image, according to classes defined by a specified taxonomy" ([Castelli et al. 1996], page 2199). At each step of the classification, the algorithm decided the class label and assigned the label to the whole block at a low level in order to be reexamined at a higher level. If at anytime the whole block was found to be homogenous, no further detail examination was required. Castelli et al. [1996] presented a wavelet-based recursive classification algorithm for progressively classifying images, which allowed the classification result to be available at each step [Castelli and Kontoyiannis 1996, 1999; Castelli et al. 1996] providing a large speedup (three to four times) in classifying large images as opposed to a pixel-by-pixel approach, and the improvement allowed by DWT was used in the application of landcover classification in image databases.

As another example of the multiscale classification, Chang and Kuo [1993] created a tree-structured wavelet transform for texture classification. To perform the classification, dominant features were chosen, which were determined by wavelet coefficients with large energy values. By repeating these wavelet decomposition and classification steps, further zoom-in into the classification of the desired frequency was possible.

In summary, signals can be decomposed into different scales of subsignals with DWT's ability to isolate signal components. Suitable time series data for wavelet-based classification have multiscale signal components that are more meaningful in parts than in sum, such as audio signals, and patients' ECG heart rates. DWT is also a suitable noise filtering technique for a preprocessing step that is necessary before performing classification. Therefore, time series data prone to noise is very suitable to be transformed with DWT before the classification. In other cases where noise is not a problem, DWT is beneficial in isolating signals into several components and performing a stepwise classification on each of these components. This last benefit is especially important in the image classification and audio classification domains, where the data has too many fine details that can be excluded from the classification process [Castelli and Kontoyiannis 1996, 1999; Castelli et al. 1996; Chang and Kuo 1993]. Classification can be performed correctly on dominant features, and even on more detailed features upon users' requests.

4.3. Wavelet-Based Clustering in Time Series

The goal of time series clustering is to group similar time series data together into the same clusters and put dissimilar time series into different clusters. Examples of applications that benefit from time series clustering include clustering patients into groups based on their clinical measurements over time, or grouping stocks in the stock market according to their price fluctuations. Clustering allows users to identify patterns and trends pertaining to each group. Time series clustering has been applied in diverse application domains, such as clustering stocks in the stock market [Fu et al. 2001; Gavrilov et al. 2000], clustering gene expressions [Balasubramaniyan et al. 2005], clustering hot and cold air pockets to predict the weather temperature [Sarma 2006], and clustering images for image retrieval [Ardizzoni et al. 1999; Natsev et al. 1999; Sheikholeslami et al. 1999]. Due to the high dimensionality, the execution of clustering algorithms is costly in terms of computational time. Research work on clustering time series data relies on transforming raw time series data using some dimensionality reduction techniques [Gavrilov et al. 2000; Lin et al. 2004], and then performing clustering on the transformed data, or on clustering the large amount of time series in small pieces recursively to improve the efficiency of the clustering process [Chaovalit 2009; Chaovalit and Gangopadhyay 2009].

There are various clustering algorithms proposed in the literature, such as k-means, CLARANS, BIRCH, DBSCAN, STING, and CLUDIS [Han and Kamber 2006; Korn et al. 1997]. Though many of these have been proposed, fast clustering algorithms typically still have a high computational complexity of $O(n^2)$ or $O(n \log n)$, where n is the number of data instances [Korn et al. 1997].

Clustering methods are evaluated for their efficiency measured by the computational complexity of different clustering algorithms. However, it is a challenge to evaluate the accuracy of the clustering methods. Since the labels of clusters are not predefined prior to clustering, data miners do not know whether the clusters are assigned to time series correctly. In order to evaluate the clustering methods' accuracy, researchers measure how well the clustering results conform to the clustering objective function, that is, distances among data within the same clusters are small, while distances among data from different clusters are large [Han and Kamber 2006]. One such evaluation metric is the Silhouette function, which is defined as the ratio of the difference between the average intracluster and the average intercluster distances, to the maximum of the intracluster and intercluster distances [Kaufman and Rousseeuw 1990]. Some articles on clustering time series evaluate their clustering by the sum square (SSQ) distance of each time series to its respective cluster centroids [Aggarwal et al. 2003; Guha et al. 2003]. A small SSQ value indicates a better cluster formation.

The use of DWT in clustering can be found in several articles [Chaovalit and Gangopadhyay 2007; Lin et al. 2004; Sheikholeslami et al. 1998]. First, WaveCluster [Sheikholeslami et al. 1998], a well-known clustering algorithm, applies DWT to filter numerical data prior to clustering. WaveCluster uses hat-shaped filters as they are best for sharpening cluster edges. They accentuate boundaries by suppressing low-frequency parts of the signal (clusters) and emphasizing high-frequency parts of the signal (boundaries). Using these filters, WaveCluster can also eliminate outliers, which is useful for clustering noisy data. WaveCluster inherits the multiresolution advantage from DWT; hence data can be clustered on various resolutions. The algorithm produces high-quality clusters especially with formations that are arbitrarily shaped.

Another example of time series clustering using DWT appears in Lin et al. [2004]. Lin et al. [2004] exploited DWT to extend the *k*-means algorithm. It is well known that one of the *k*-mean's disadvantages is random seed selection which results in different clusters when executed multiple times. Therefore, cluster qualities are highly dependent on initial seeds. DWT can alleviate this problem using its multiresolution property. The authors proposed Iterative *k*-means algorithm (*I-k*means) which has the following steps. First, DWT is applied onto time series. Then, the *k*-means algorithm is executed on the coarsest level of decomposition in order to obtain cluster centers. Subsequently, the DWT coefficients are reconstructed to the next higher level of decomposition. For each next level of decomposition, the cluster centers which were previously obtained from clustering the prior level of decomposition are used as seeds. In clustering the newly reconstructed DWT coefficients at finer levels of decomposition, good initial seeds have been selected from the approximation since the very beginning of the clustering process. The *I-k*means algorithm iterates until the finest level of decomposition but can be stopped at any decomposition level by users. Clustering time series in this manner eliminates cluster centers falling into local minima, hence producing better results than the traditional *k*-means algorithm. The *I-k*means algorithm has an advantage of better clustering qualities and faster processing, which are direct benefits from initializing seeds at the approximation level of time series. As the number of dimensions of time series is reduced at lower levels of decomposition, *I-k*means saves time of clustering on full-resolution data by doing most of the work early on when the number of time dimensions is low.

In addition to the above mentioned works, other researchers have applied DWT on clustering for different purposes [Cheong et al. 2005; Ghosh-Dastidar and Adeli 2003; Li et al. 2000a]. The multiresolution property of DWT, which is the most unique property among data dimensionality reduction techniques in the literature, has been useful in clustering. For example, Li et al. [2000a] and Cheong et al. [2005] analyzed air pollutant data and stock data, respectively, using multiresolution DWT coefficients. The multiresolution property allows the underlying trends and localized patterns found in wavelet coefficients to be analyzed. In Cheong et al. [2005], the decomposed time series of stock data were clustered for similarities among stocks. It was found that groups of stock at different temporal resolutions corresponded to real-world events. As an example, the connection between two stocks with an owner-subsidiary relationship was found when data was decomposed. This connection was not discovered from clustering raw time series data due to noise. While a property such as the multiresolution capability was more directly exploited, noise reduction was sometimes a byproduct. DWT's noise reduction was intentional in situations where high signal-to-noise ratios were present [Wang et al. 2006]. In Wang et al.'s [2006] work, hyperspectral Raman data have high signal-to-noise ratio which requires an effective preprocessing in order to cluster chemical groups of this data. Although conventional smoothing methods can achieve the purpose of noise reduction, they will distort important features. DWT was chosen by Wang et al. [2006] because it was able to distinguish between features and noise. Compared to other denoising approaches (Spline filter and Savittzky-Golay), DWT was able to smooth data while producing higher levels of clustering accuracy when used by the fuzzy c-means clustering.

Although we have discussed the advantages that DWT offers on clustering, the benefits of DWT in clustering was sometimes more secondary. For example, Sheikholeslami et al. [1999] employed DWT as part of an image retrieval system because clustering on an image database allows faster image retrieval and better accuracy. DWT was used to control the level of details to compare and cluster images for further retrieval.

DWT was not used only with Euclidean distance. An example of applying other distance functions with DWT can be found in Ghosh-Dastidar and Adeli [2003], employed the Mahalanobis distance when clustering traffic data with wavelets. Traffic data was represented as speed, volume, and occupancy. Therefore, the Mahalanobis distance was chosen in order to capture correlations among these components in traffic data. Also, the authors investigated the use of several wavelet filtering schemes which were implied from the domain knowledge. The different wavelets included in their study were Haar, second-order Daub, second-order Coifman, and fourth-order Coifman. These wavelets' coefficients were selected and applied to neural networks for clustering. The results showed that the wavelet-based clustering technique proposed by Ghosh-Dastidar and Adeli [2003] gave the most superior results when the fourth-order Coifman wavelets were employed.

4.4. Multiresolution Anomaly Detection

Anomaly detection deals with detecting anomalous behavior of time series data. It helps users identify abnormal time series or parts of time series from the rest of the data. The concept of anomaly detection considers the behavior from past data and/or data models to perform this task. Any data that deviates from its regularity is considered an anomaly. Anomaly detection is critical in various application domains, such as intrusion detection in computer networks. One main challenge in anomaly detection is how to successfully distinguish an anomaly.

Usually, an *anomaly* refers to an individual data point or a set of data points that deviate from an expectation. Sometimes, the concept of an anomaly is interchangeable with the concepts of a surprise or interestingness. We usually hear the terms *surprise detection* [Shahabi et al. 2000, 2001], "*event detection* [Atallah et al. 2004; Bunke and Kraetzl 2004; Guralnik and Srivastava 1999; Saeed and Mark 2001], *change detection* [Zeira et al. 2004], "*burst detection* [Shasha and Zhu 2004], and "*novelty detection* [Dasgupta and Forrest 1995; Keogh et al. 2002; Ma and Perkins 2003; Marsland 2001] in the anomaly detection literature [Bailey et al. 1998; Chin et al. 2005; Klimenko et al. 2002; Lane and Brodley 1999; Lee and Stolfo 1998; Luo et al. 2001; Magnaghi et al. 2004; Mallat and Hwang 1992; Shmueli 2004; Struzik and Siebes 2000; Wang 1995]. Yet, it is not surprising that these various terms may refer to the same general concept.

Let us discuss some of the various definitions for anomaly detection that have been proposed in the literature. First of all, the concept of *outlier detection* comes to mind. An outlier is a data point in a time series that differs greatly from other data points [Keogh et al. 2002]. Therefore, outlier detection involves a problem of finding data points that diverge from the normal range. Another term in the literature with the same definition as outlier detection is *deviation detection* [Arning et al. 1996]. In sum, outliers are a set of potentially interesting data points, since they depart significantly from the expectation, and should be subjected to further investigation.

Next, let us examine other related concepts. Shahabi et al. [2000] defined "surprises" as sudden changes within the time series. Bunke and Kraetzl [2004] detected an "abnormal change" when a similarity between two time series of graphs fell below a certain threshold. Keogh et al. [2002] defined *surprising patterns* as a collection of data points whose "occurrence frequency differ substantially from that expected by chance, given some previously seen data. In fact, a surprise could also be a concept among various types of interestingness [Geng and Hamilton 2006].

When data points are considered collectively and not individually, the problem has shifted from outlier detection to *anomaly detection*. For anomaly detection, the goal is to find interesting data patterns that are surprising or unexpected.

Various domain applications have utilized the anomaly detection task and its relatives. Detection of abnormal events includes network performance problem detection [Huang et al. 2001], intrusion detection [Lee and Stolfo 1998; Luo et al. 2001], disease outbreak detection [Wong 2004], bio-terrorist attack detection [Shmueli 2004], and change detection in classification models [Zeira et al. 2004]. Since all of the terms mentioned above are subjective and application-dependent, it is crucial for researchers to have a clear definition of their "anomaly" for researching anomaly detection.

Given a variety of definitions on the concept, the approaches to anomaly detection can vary. These approaches include finding dramatic shifts in the time series [Shahabi et al. 2000], change point detection [Guralnik and Srivastava 1999], comparing a new time series to reference time series or patterns [Chen et al. 1999a, 1999b], and discriminating time series between self and nonself [Dasgupta and Forrest 1995; Forrest et al. 1994]. Work in the field has introduced various ways to identify an anomaly, for example, by utilizing probabilities [Atallah et al. 2004; Chin et al. 2005; Guralnik and Srivastava 1999; Keogh et al. 2002; Wong 2004], similarity measures [Arning et al. 1996; Keogh et al. 2004b; Lane and Brodley 1999; Wei et al. 2005a, 2005b], rule induction [Keogh et al. 2002; Luo et al. 2001], matching functions [Forrest et al. 1994; Ma and Perkins 2003], and graph-based edit distance [Bunke and Kraetzl 2004]. We can group anomaly detection approaches into three categories.

(1) detecting abrupt changes within a time series [Fu et al. 2006; Guralnik and Srivastava 1999; Shahabi et al. 2000];
(2) detecting abnormality by comparing distances between two or more time series [Arning et al. 1996; Bunke and Kraetzl 2004; Wei et al. 2005a, 2005b];
(3) detecting irregular patterns by observing the regular frequency of data points [Atallah et al. 2004; Chin et al. 2005; Keogh et al. 2002; Lee and Stolfo 1998; Luo et al. 2001; Wong 2004].

The approaches in the first category do not need previous data or other time series data for comparison. On the contrary, the approaches in the second and third categories need at least two time series to be present. However, if only a single time series is available, such detection is achievable by applying a sliding window to segment the time series. Then one can specify some of these segmented time windows as a reference segment [Fu et al. 2006; Wei et al. 2005b].

As an example, one study at the University of California, Riverside, utilized the combination of two techniques: Chaos Game Representations (CGR) and Symbolic Aggregate approXimation (SAX) data representation [Wei et al. 2005a, 2005b]. SAX is a symbolic time series data representation, which is useful for converting real-value sequences into discrete data. In Wei et al. [2005a], the CGR technique was used to map sequences of discrete values, which were previously transformed by SAX, into a $2L$ by $2L$ grid bitmap, where L is the length of the sequences. Subsequently, the frequency of pixels within the grid bitmap was counted and color-coded to allow the human eye to compare and contrast. A $2L \times 2L$ grid bitmap represented a single, sequence and color-coding helped researchers distinguish among various sequences.

Wei et al. [2005b] focused on anomaly detection. The basic idea was to create two concatenated windows and to slide them together across the sequence. The anomaly score for each pair of two sliding windows was plotted along the time scale. According to the experiments conducted in both studies [Wei et al. 2005a, 2005b], an anomaly can be detected via visualization. While the first study emphasized visualizing each time series, the second approached anomaly detection by visualizing the difference between two time series.

Most of the standard anomaly detection approaches (perhaps with the exception of Wei et al.'s [2005a, 2005b] work) cannot detect anomalies at different temporal scales, while wavelet-based approaches can. In addition, wavelet-based anomaly detection approaches offer the advantage of selecting the levels of resolution to detect anomalies from. This sets the wavelet-based approaches apart from Wei et al.'s [2005a, 2005b] approach. While the latter performed discretization on time series at different scales and applied colored visualization on the discretized values, the former can combine some different scales of time series wavelet coefficients while intentionally skipping others. This benefit was demonstrated in Shahabi et al. [2000] when only some coefficients of time series were retained due to space limitation. Shahabi et al. [2000] proposed various strategies of coefficient dropping, but suggested that selective coefficient dropping was appropriate when time series contain many outliers. The selective dropping of coefficients proved to be fitting for the surprise detection purpose, because dropping coefficients means abandoning some unnecessary information in the time series for anomaly detection. This ability is not found in the standard anomaly detection approaches. Even with other data reduction techniques such as SVD and DFT, time series data can be reduced to coefficients but the anomaly information is not isolated as well as in the case of DWT.

Due to its ability to separate original time series into its decompositions, DWT is a powerful tool to help researchers capture trends, surprises, and patterns in data. It is also the data transformation technique that concurrently localizes both time and frequency information from the original data in its multiscale representation. Other techniques, such as discrete Fourier transform (DFT) and discrete cosine transform (DCT), convert data from the time domain into the frequency domain, but in doing so temporal semantics—the sense of when significant events happen—is lost. In contrast, auto regressive moving average (ARMA) models preserve the temporal information in their results, but lose the frequency information [Dillard and Shmueli 2004]. In other words, techniques such as DFT and DCT do not have a time localization property, while ARMA models do not have a scale localization property [Dillard and Shmueli 2004].

The scale localization property of DWT makes the anomaly detection task at different resolutions both feasible and promising. A series of wavelet approximation coefficients at the ith level ($A_{ij} = \{A_{i1}, A_{i2}, \ldots, A_{ik}, j = 1 \ldots k$) represents trends, while a series of wavelet detail coefficients at the ith level ($D_{ij} = \{D_{i1}, D_{i2}, \ldots, D_{ik}\}, j = 1 \ldots k$) represents surprises for that scale. Repeated patterns of signals can be discovered in both A_{ij} and D_{ij}—the products of wavelet transform. More importantly, trends, surprises, and repeated patterns identified by DWT also preserve temporal semantics.

Studies that have utilized DWT for anomaly detection always draw on the visualization of wavelet-transformed data as a part of their approach [Dillard and Shmueli 2004; Huang et al. 2001; Shahabi et al. 2000]. In time series data, visualizing time series data is a preliminary technique for descriptive analysis. Assuming that the signals are nonstationary with several mixed components, DWT reveals those patterns that are hidden in the original signals. Visualizing wavelet-transformed signals helps illuminate the trends in approximations, surprises in details, and repeated patterns in different levels of decomposition.

Figure 6 (left-hand side) visualizes decompositions where the trends of this signal have been enhanced. The higher the level of resolution, the smoother the trend we perceive. Such a trend pattern is harder to pick out in raw time series data, especially when the time series has a lot of spikes and peaks. Also an on the right-hand side of the

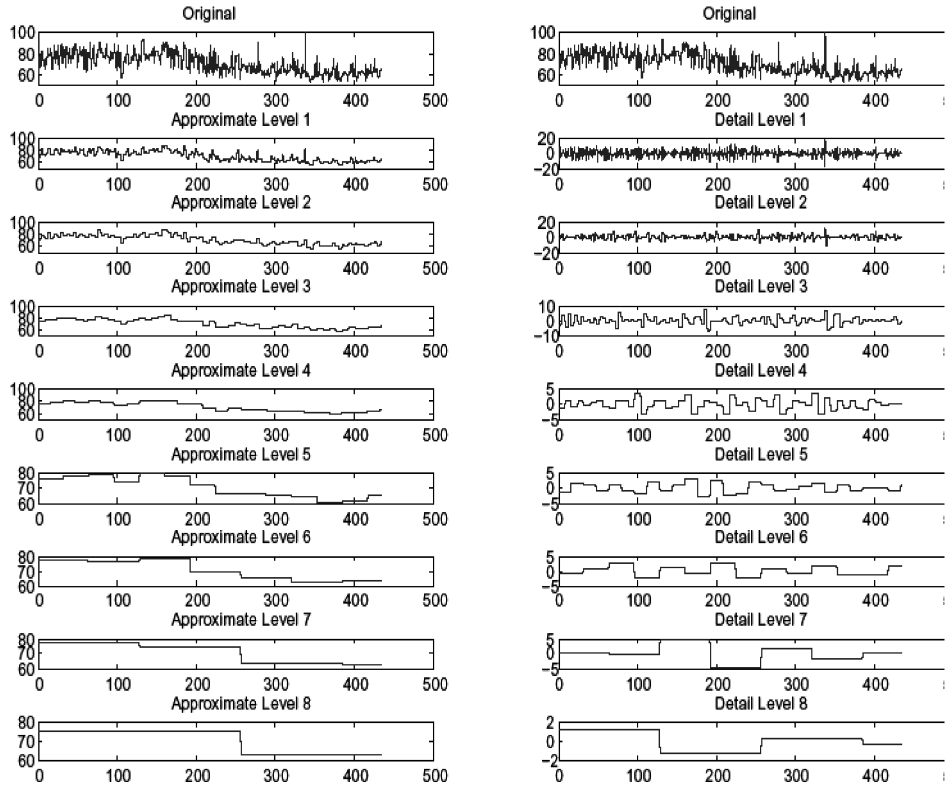

Fig. 6. A visualization of trends, surprises, and patterns at different decomposition levels by Haar wavelet transform.

same figure, surprises have been filtered out from the signals and appear in the detail graphs. The surprises are distinguishable at different levels of decomposition. Each surprise is at a different scale and at a different time location. Therefore, surprises are a good indicator of singularities in signals. The singularities are preserved in both scale and time location in the wavelet decomposition.

In applying DWT to time series, peaks or spikes that are in one scale may not be as obvious in another scale. The scale localization property enables wavelets to capture details pertaining to such scales and such scales only. Suppose that, for decompositions D1–D4, where D1 corresponds to details at a 1-day interval, D2 at a 2-day interval, D3 at a 4-day interval, and D4 at an 8-day interval, spikes that are visible at $t = 100$ for D1 and D2 might not be apparent at D3 or beyond. Those spikes imply that a surprise occurs at $t = 100$ for D1 and D2, for both 1-day and 2-day intervals, and that this particular surprise does not pertain to larger time scales. Conversely, suppose that there are some surprises in the original data showing at D4, but not at D3 or other lower levels of decomposition. Consequently, these surprises correspond to events in the 8-day interval, but not in the smaller or larger time scales.

Looking closely at the detail graphs in Figure 6, we see repeated patterns of spikes appearing in some levels of detail coefficients. A zoom-in version of these graphs is shown in Figure 7. By interpreting the visualization in this manner, experts with domain knowledge can provide better insights into the semantics of trends, surprises, and repeated patterns. As noted before, this type of surprise information cannot be captured with other time series analysis techniques [Dillard and Shmueli 2004; Mallat

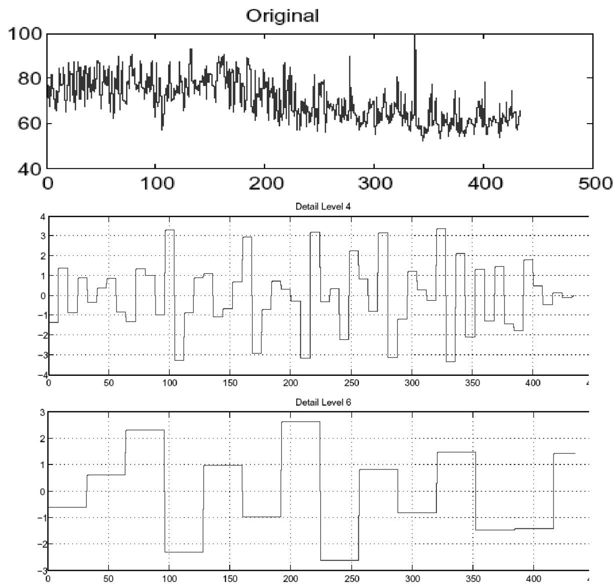

Fig. 7. A visualization of (a) a zoomed-in original signal (top), (b) repeated patterns in D4 (middle), and (c) repeated patterns in D6 (bottom).

and Hwang 1992]. However, the analysis methodology of the wavelet-based anomaly detection has been mostly limited to visualization [Dillard and Shmueli 2004; Shahabi et al. 2000].

Besides using wavelet coefficients and wavelet coefficient graph, wavelet-based surprise detection can be used to create another type of graph: an energy plot. Previous studies on anomaly detection in time series data utilized an energy plot as a tool to detect irregular patterns [Huang et al. 2001; Magnaghi et al. 2004]. This approach was more quantitative in nature than the visualization approach. In these studies, a wavelet energy function at the qth level of decomposition among all levels was defined as

$$E_q = \frac{1}{N_q} \sum_k |d_{q,k}|^2,$$

where E_q denotes an energy at level q, N_q denotes the number of coefficients at scale q, and $d_{q,k}$ denotes a detail wavelet coefficient at a position k of level q [Huang et al. 2001]. The scale q was plotted along the x axis, and a logarithmic energy at scale q, $log_2(E_q)$ was plotted along the y axis. By knowing that the logarithmic values of energy function would remain constant for white noise time series [Huang et al. 2001], any apparent "dips" in the energy plot illustrated low values in the energy function, which in turn indicated irregular events in time series. Therefore, the logarithmic energy plot could show a relationship between the scale q and the logarithmic energy at scale q, $log_2(E_q)$, for time series and illustrate any surprise in data [Huang et al. 2001; Magnaghi et al. 2004].

Other studies employed different techniques for detecting energy dips in the energy plot. By using energy plots, Magnaghi et al. [2004] detected the local minima of the energy function in a least-square parabola. Huang et al. [2001] visually compared the energy plot with a round-trip time (RTT) plot and a retransmission timeout (RTO) plot to assure the alignment of energy dips among the corresponding measurements. Energy plots of the investigated time series were then compared to a baseline energy plot, which illustrated a normal condition of the computer network path. If the difference between these two plots was greater than a threshold δ, an irregular pattern was detected. Such surprises are visible through dips in the energy plots. The experiments presented promising results in Huang et al. [2001] and Magnaghi et al. [2004]. However, these studies utilized the scale-localization property but did not reach as far as to analyze the timing of the anomaly. Generally speaking, this approach has not yet quantitatively utilized the time-localization property.

DWT has also been employed for detecting anomaly in time series as a discretization technique [Fu et al. 2006]. In Fu et al.'s [2006] approach, subsequences of time series were transformed using Haar wavelets and then the coefficients were converted into symbols. These strings of symbols formed words. The technique presented by Fu et al. [2006] exploits DWT to adjust an effective word length in order to compress the time series subsequences. Results showed that DWT reduced the number of times the distance function was called, when compared to the baseline algorithm without DWT.

The application of DWT in anomaly detection has appeared in diverse application domains, such as manufacturing [Li et al. 1999, 2000b; Yao et al. 1999], disease outbreak detection [Shmueli 2004; Wong 2004], and anomalies in computer networks [Huang et al. 2001; Magnaghi et al. 2004]. We find that these approaches do not yet fully utilize the benefits of DWT's multiresolution analysis. Multiresolution analysis can be used to further analyze data in both time and scale correspondences, and this capability has not been shown or implemented among the studies that we reviewed. With the combination of both time- and scale-localization properties, DWT has more potential than is currently being exploited. And, therefore, opportunities for analyzing time series for trend, surprise, and pattern detection through DWT's multiresolution analysis are enormous.

4.5. Wavelet-Based Prediction

Time series data prediction is the task of forecasting data values in the future, given input values such as a prediction time point and historical data values. In general, people are very interested in forecasting time series data, such as stock prices [Fu et al. 2001; Gavrilov et al. 2000], weather [Sarma 2006], electricity and water consumption [Collin 2004; Dillard and Shmueli 2004; Petridis et al. 2001], river water level, Internet usage [Basu et al. 1996], disease outbreak [Banner et al. 1976; Goodwin and Maher 2000; Shmueli 2004; Silver and Ginsburg 1984; Wong 2004], and physiological symptoms [Banner et al. 1976; Goodwin and Maher 2000; Shmueli 2004; Silver and Ginsburg 1984; Wong 2004], to name a few.

Prediction or forecasting is considered as a fitting process of time series data to a model [Anderson 1976, 1997; Brockwell and Davis 1991]. All predictive models, such as those weather forecasting and stock prediction, need historical data as input for developing such a model. The time series prediction process starts with developing a mathematical hypothesis of a model that represents the input data [Brockwell and Davis 1991]. Factors hypothesized in the model are those that affect the values of time series. Then parameters are derived from these factors, and are estimated using an appropriate time series analysis technique. As a general rule, the larger the available time series data collection, the better the model and parameters estimation is, and hence the more precise the prediction becomes. Next the model is evaluated using a goodness-of-fit test, which indicates how robust the model is for time series data. In a general sense, a goodness-of-fit test returns errors between predicted data and actual data. If the model verifiably describes the underlying data well (with few errors), future values of these observations can be predicted using this model, given the assumption that the behavior of future data remains constant. If the fitness of the model is not yet satisfied, the model needs to be adjusted and reverified. Once the goodness of fit of the model is satisfied for a particular set of time series, the model is ready for use with new time series datasets.

One may derive data characteristics of the time series from the model fitting process. For example, if a time series is generated from a stochastic process, where the probability distribution of the data is fixed along the time axis, the characteristics of that signal—namely, a stationary signal—will have a mean and a variance that do not fluctuate over time [Anderson 1976]. Such a characteristic may be interpreted from looking at the model. For time series prediction, a time series might have combined characteristics of various types, for example, trend, seasonal, and noise, and the study of time series is about the discovery of those characteristics. The main objective is to find a mathematical model which accurately describes those characteristics in order to represent that particular set of time series. When the model is discovered, it can also be used in several applications beyond prediction, including noise filtering and future value control [Brockwell and Davis 1991].

As with other time series data analysis and mining approaches, researchers need to evaluate time series data prediction. An error measurement is usually employed to mark the quality of the prediction. A popular error measurement is the *root mean square error* [Korn et al. 1997; Weigend and Gershenfeld 1994]. At times, this measurement will be normalized with its respective standard deviation to obtain a more accurate evaluation across different data. Root mean squared percent error (RMSPE) was defined by Korn et al. [1997].

In brief, the main idea behind time series prediction is to understand the movement of historical data and apply this understanding for future prediction. Such movements, when analyzed by DWT, display important patterns more obviously, thus enabling researchers to perform a prediction task more effectively. Researchers have modeled time series into trend, seasonal, and noise components. The models can be constructed in various ways depending on an underlying assumption of data. This section discusses a number of studies that utilize DWT for time series prediction.

A group of researchers (Murtagh, Starck, and Renaud) focused their work on *à trous wavelet transform* [Murtagh et al. 2004; Renaud et al. 2003, 2005]. *À* trous wavelet transform is a redundant form of the regular DWT. *À* trous coefficients are created by shifting wavelet functions on time series data one point at a time, instead of 2^j points at a time as with the normal DWT, where j denotes the current level of resolution.

In Murtagh et al. [2004] and Renaud et al. [2003, 2005], *à* trous is advantageous for wavelet-based prediction since it allows a one-step forward prediction and avoids the problem of finding corresponding coefficients with the original time window for prediction, also known as the *boundary problem* in the wavelet literature. Wavelet coefficients are selected from each scale to perform a multiscale prediction. Experimentation has shown that these methods have superior results on two prediction schemes: autoregressive (AR) models and neural networks. The authors also claim that the proposed approach is easily extendable to other prediction schemes as well.

Lotrič [2004] and Soltani et al. [2000] performed predictions based on a similar concept of multiscale wavelet decomposition. The difference among these studies is that the work of Soltani et al. [2000] chose all the coefficients of *à* trous wavelet transform, while Lotrič [2004] and Murtagh et al. [2000] chose a particular set of coefficients according to scales [Renaud et al. 2005].

In summary, studies on wavelet-based prediction have explored and utilized the multiresolution property of DWT and the *à* trous wavelet transform. For the *à* trous wavelet transform, predicting the next values of a time series requires that none of the corresponding wavelet coefficients be calculated from unknown upcoming data points. Hence, the *à* trous wavelet transform has been applied due to the application of time series prediction.

5. CONCLUSION

In this article, we reviewed the literature in the field of discrete wavelet transform (DWT), and the application of DWT on time series data analysis and mining. We have also illustrated the potential of applying DWT on time series data analysis and mining, especially its multiresolution analysis. A large number of studies demonstrated the applicability of DWT to data analysis and mining for various domain applications. We found that many desirable properties of DWT have been realized and practiced by research communities.

Researchers have used DWT for noise reduction in various domain-specific data such as audio data for a better audio classification system, and medical data for better illness diagnosis. In addition, DWT is an effective dimensionality reduction technique to apply before conducting a similarity search. It greatly reduces search time, while preserving accuracy. DWT is unique for its multiresolution analysis, which allows researchers to apply it at different levels of data resolution, resulting in significant benefits such as faster data mining process, less data storage, and better mining results. Domain applications that benefit from applying DWT to time series data analysis include, but are not limited to, image querying, audio querying, illness classification, image texture classification, satellite image classification, pornographic image classification, audio and video genre classification, computer network anomaly detection, and disease outbreak detection. DWTs are useful in defining new sets of features used in classification and similarity search applications from wavelet coefficients. These usually result in more well-defined features due to the reduction in noise or irrelevant data, which in turn increase the accuracy of classification and similarity search. Another apparent use of DWTs is allowing researchers the freedom to investigate data at different temporal scales. For example, patterns of wavelet coefficient energy at different scales have been used to detect network anomaly. Other studies take this benefit further to perform a progressive time series analysis, such as progressive classification. Progressive time series analysis is beneficial in such a way that it is a step-by-step approach, in which the first few steps allow researchers to mine data at coarse levels, producing somewhat approximate answers while reducing the processing time (they can perform additional steps for finer data if they need to).

The research included in this survey mostly employed a limited number of wavelet filters and distance functions. Frequently, the Haar filter and the Euclidean distance are used. Nevertheless, the lack of diversity in wavelet filters and distance functions does not indicate a limitation of DWT in these areas. In general, DWT can handle other distance functions besides Euclidean, and different wavelet filters have been applied to time series analysis. An example of such varieties can be found in Ghosh-Dastidar and Adeli [2003], who utilized a different distance function (Mahalanobis distance) and various wavelet filters in order to search for the most appropriate wavelet filters for analyzing their traffic data. As another example, Coifman and Wickerhauser [1992] used cost functions such as Shannon entropy to select the best basis functions to the given signal.

Before employing DWT, however, there are some related challenges that one needs to address. These challenges relate to each of the following issues.

(1) *Choice of wavelets*. An appropriate wavelet filter can be identified, as illustrated in Ghosh-Dastidar and Adeli [2003] and Sheikholeslami et al. [2000], wheredifferent filters were compared or a special property of a particular wavelet filter was exploited. In that case, other wavelet filters may be found more appropriate than the simple Haar wavelet. Nevertheless, Haar has often been found an appropriate filter in various research studies, such as Chan and Fu [1999].
(2) *Depth of analysis*. This issue deals with the number of levels of decomposition. It is theoretically possible to decompose data up to the coarsest level, but at each

level the approximate data is being filtered. How does one know when to stop the decomposition? At which level of decomposition should the signals be analyzed? One research study [Lalitha 2004] answered this question by measuring the entropy of wavelet coefficients at each level and finishing the decomposition when a stopping condition was met. A heuristics applied by Lalitha [2004] prevented the interesting trend of degrading fault signals in gas turbines data from being further distorted by identifying the optimal level of decomposition.

(3) *Boundary problem*. Computation of wavelet coefficients for a given level of decomposition requires a certain number of data samples. In real-life situations, it is possible for a dataset to contain insufficient number of samples for calculation. This may happen when (i) data is irregularly sampled, (ii) some data observation values are missing, and (iii) the number of samples is not enough for computation. This problem is referred to as the *boundary problem* and has no universally best solution. One must make an informed decision based on the advantages and disadvantages of the following boundary correction methods.

There are solutions to the boundary problem proposed by Jensen and Cour-Harbo [2001] and Ogden [1997]. They either applied treatments to data or wavelet filters to solve the problem. The simplest solution to implement is the zero-padding technique, where missing samples are added as zeros into the data sequence. In this case, the manipulated data variances are tampered and the orthogonality of data is not preserved. Other methods from Ogden [1997] include data interpolation and numerical integration. The former requires creating a new dataset through interpolation of the original dataset in order to create good approximations of wavelet coefficients without changing data distribution. However, this approach introduces some amount of correlations among coefficients. The latter employs numerical integration in computing top-level wavelet coefficients. This approach introduces the least amount of artificial data into the coefficients but is computationally expensive. More frequently used methods are available for further reading in Jensen and Cour-Harbo [2001], which include boundary filters, periodization, and mirroring. In boundary filters, new filter coefficients at each end of the signal are substituted in order to preserve the perfect reconstruction of data without modifying the signal's lengths. Periodization chooses samples from the signal to add into the sequence instead of zero padding. Mirroring first mirrors a signal and then adjoins the result to the original signal. Periodization is then performed on the adjoined data to truncate the signal. Both periodization and mirroring can lead to incorrect wavelet analysis as discontinuities may be present from truncating data samples. However, mirroring is popular in image applications, where symmetry is preferred to the eye.

(4) *Data dependency*. DWT is a data-dependent technique. As pointed out by Shasha and Zhu [2004], DWT requires time series data to have principal components. When data is stationary and/or when patterns do not exist in data, DWT may not necessarily be superior to other methods.

Perhaps one area where DWT has not been fully utilized in the literature so far is taking temporal semantics of wavelet coefficients into account when performing data analysis. DWTs have been realized for its multiresolution capability, but only in a relative sense to another level. A limited number of works have derived meaningful temporal semantics, that is, in absolute time scales such as weekly or monthly patterns, from the mining results. Moreover, anomaly detection or surprise detection using DWT is still largely done with visualization. Since experts usually take a look at plots of wavelet detail coefficients or plots of other types to detect anomalies, model-based anomaly detection such as those in nonwavelet time series anomaly detection has yet to be formalized.

REFERENCES

AGGARWAL, C. C., HAN, J., WANG, J., AND YU, P. S. 2003. A framework for clustering evolving data streams. In *Proceedings of the 29th International Conference on Very Large Data Bases (VLDB)*.

AGRAWAL, R., FALOUTSOS, C., AND SWAMI, A. 1993. Efficient similarity search in sequence databases. In *Proceedings of the 4th International Conference on Foundations of Data Organization and Algorithms (FODO)*. 69–84.

ANDERSON, J. G. 1997. Clearing the way for physician's use of clinical information systems. *Comm. ACM 40*, 83–90.

ANDERSON, O. D. 1976. *Time Series and Forecasting: The Box-Jenkins Approach*. Butterworths, London, U.K.

ARDIZZONI, S., BARTOLINI, I., AND PATELLA, M. 1999. Windsurf: region-based image retrieval using wavelets. In *Proceedings of the 10th International Workshop on Database and Expert Systems Applications (DEXA)*. 167–173.

ARNING, A., AGRAWAL, R., AND RAGHAVAN, P. 1996. A linear method for deviation detection in large databases. In *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (SIGKDD)*. 164–169.

ATALLAH, M., GWADERA, R., AND SZPANKOWSKI, W. 2004. Detection of significant sets of episodes in event sequences. In *Proceedings of the 4th IEEE International Conference on Data Mining (ICDM)*. 3–10.

BAILEY, T. C., SAPATINAS, T., POWELL, K. J., AND KRZANOWSKI, W. J. 1998. Signal detection in underwater sounds using wavelets. *J Amer. Statist. Ass. 93*, 441, 73–83.

BALASUBRAMANIYAN, R., HÜLLERMEIER, E., WESKAMP, N., AND KÄMPER, J. 2005. Clustering of gene expression data using a local shape-based similarity measure. *Bioinformatics 21*, 7, 1069–1077.

BANNER, A. S., SHAH, R. S., AND ADDINGTON, W. W. 1976. Rapid prediction of need for hospitalization in acute asthma. *J. Amer. Medical Ass. 235*, 13, 1337–1338.

BASU, S., MUKHERJEE, A., AND KLIVANSKY, S. 1996. Time series models for Internet traffic. In *Proceedings of the 15th Annual Joint Conference of the IEEE Computer and Communications Societies, Networking the Next Generation (INFOCOM96)*. 611–620.

BLUME, M., AND BALLARD, D. R. 1997a. Image annotation based on learning vector quantization and localized Haar wavelet transform features. In *Proceedings of the Applications and Science of Neural Networks Conference (SPIE)*. 181–190.

BLUME, M. AND BALLARD, D. R. 1997b. Image annotation based on learning vector quantization and localized Haar wavelet transform features. In *Proceedings of the Applications and Science of Artificial Neural Networks III Conference (SPIE)*, S. K. Rogers, Ed. 181–190.

BRAMBILLA, C., VENTURA, A. D., GAGLIARDI, I., AND SCHETTINI, R. 1999. Multiresolution wavelet transform and supervised learning for content based image retrieval. In *Proceedings of the IEEE International Conference on Multimedia Computing and Systems (ICMCS)*. 9183–9188.

BROCKWELL, P. J. AND DAVIS, R. A. 1991. *Time Series: Theory and Methods*. Springer-Verlag, Berlin, Germany.

BUNKE, H. AND KRAETZL, M. 2004. Classification and detection of abnormal events in time series of graphs. In *Data Mining in Time Series Database*. World Scientific Publishing, Singapore, 127–148.

CASTELLI, V. AND KONTOYIANNIS, I. 1996. Wavelet-based classification: Theoretical analysis. Tech. rep. RC-20475, IBM T. J. Watson Research Center, Yorktown Heights, NY, 1–25.

CASTELLI, V. AND KONTOYIANNIS, I. 1999. An efficient recursive partitioning algorithm for classification, using wavelets. Tec. rep. RC-21039, IBM T. J. Watson Research Center, Yorktown Heights, NY, 1–27.

CASTELLI, V., LI, C.-S., TUREK, J., AND KONTOYIANNIS, I. 1996. Progressive classification in the compressed domain for large EOS satellite databases. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2199–2202.

CHAKRABARTI, K., KEOGH, E., MEHROTRA, S., AND PAZZANI, M. 2002. Locally adaptive dimensionality reduction for indexing large time series databases. *ACM Trans. Datab. Sys. 27*, 2, 188–228.

CHAN, K.-P. AND FU, A. W.-C. 1999. Efficient time series matching by wavelets. In *Proceedings of the 15th International Conference on Data Engineering (ICDE)*. 126–133.

CHANG, T. AND KUO, C.-C. J. 1993. Texture analysis and classification with tree structured wavelet transform. *IEEE Trans. Image Process. 2*, 4, 429–441.

CHAOVALIT, P. 2009. *Clustering Trans. Data Streams by Example and by Variable*. Information Systems, University of Maryland, Baltimore County, Baltimore, MD, 203.

CHAOVALIT, P. AND GANGOPADHYAY, A. 2007. A method for clustering time series using connected components. In *Proceedings of the 17th Annual Workshop on Information Technologies and Systems (WITS)*.

CHAOVALIT, P. AND GANGOPADHYAY, A. 2009. A method for clustering transient data streams. In *Proceedings of the 24th Annual ACM Symposium on Applied Computing*. 1518–1519.

CHEN, B. H., WANG, X. Z., YANG, S. H., AND MCGREAVY, C. 1999a. Application of wavelets and neural networks to diagnostic system development, 1: Feature extraction. *Comput. Chem. Eng. 23*, 899–906.

CHEN, B. H., WANG, X. Z., YANG, S. H., AND MCGREAVY, C. 1999b. Application of wavelets and neural networks to diagnostic system development, 2: An integrated framework and its application. *Comput. Chem. Eng. 23*, 945–954.

CHEONG, C. W., LEE, W. W., AND YAHAYA, N. A. 2005. Wavelet-based temporal cluster analysis on stock time series. In *Proceedings of the International Conference on Quantitative Sciences and Its Applications (ICOQSIA)*.

CHIN, S. C., RAY, A., AND RAJAGOPALAN, V. 2005. Symbolic time series analysis for anomaly detection: A comparative evaluation. *Signal Process. 85*, 1859–1868.

COIFMAN, R. R. AND WICKERHAUSER, M. V. 1992. Entropy-based algorithms for best basis selection. *IEEE Trans. Inform. Theor. 38*, 2, 713–718.

COLLIN, N. 2004. Time-series prediction of a waste water treatment plan. D Master's thesis. Department of Numerical Analysis and Computer Science, Royal Institute of Technology, Stockholm, Sweden.

DASGUPTA, D. AND FORREST, S. 1995. Novelty detection in time series data using ideas from immunology. In *Proceedings of the International Conference on Intelligent Systems*, 1-6.

DILLARD, B. AND SHMUELI, G. 2004. Simultaneous analysis of multiple time series using two-dimensional wavelets. Department of Decision and Information Technologies, University of Maryland, College Park, MD, 1–19.

DINH, P. Q., DORAI, C., AND VENKATESH, S. 2002. Video genre categorization using audio wavelet coefficients. In *Proceedings of the 5th Asian Conference on Computer Vision (ACCV)*. 23–25.

FALOUTSOS, C., RANGANATHAN, M., AND MANOLOPOULOS, Y. 1994. Fast subsequence matching in time-series databases. In *Proceedings of the International Conference on Management of Data (SIGMOD)*. R. T. Snodgrass and M. Winslett, Eds. 419–429.

FORREST, S., PERELSON, A. S., ALLEN, L., AND CHERUKURI, R. 1994. Self-nonself discrimination in a computer. In *Proceedings of the IEEE Symposium on Research in Security and Privacy*. 1994, 202–212.

FU, A. W.-C., LEUNG, O. T.-W., KEOGH, E., AND LIN, J. 2006. Finding time series discords based on Haar transform. In *Proceedings of the 2nd International Conference on Advanced Data Mining and Applications (ADMA)*, X. Li, O. R. Zaiane, and Z. Li, Eds. Springer Berlin/Heidelberg, Germany, 31–41.

FU, T.-C., CHUNG, F.-L., NG, V., AND LUK, R. 2001. Pattern discovery from stock time series using self-organizing maps. In *Proceedings of the KDD Workshop on Temporal Data Mining*. 27–37.

GAVRILOV, M., ANGUELOV, D., INDYK, P., AND MOTWANI, R. 2000. Mining the stock market: Which measure is best? In *Proceedings of the 6th International Conference on Knowledge Discovery and Data Mining (SIGKDD)*. 487–496.

GENG, L. AND HAMILTON, H. J. 2006. Interestingness measures for data mining: A survey. *ACM Comput. Surv. 38*, 3, 1-32.

GEURTS, P. 2001. Pattern extraction for time-series classification. In *Proceedings of the 5th European Conference on Principles of Data Mining and Knowledge Discovery (PKDD)*. Springer-Verlag, Berlin Germany, 115–127.

GHOSH-DASTIDAR, S. AND ADELI, H. 2003. Wavelet-clustering-neural network model for freeway incident detection. *Comput. Aid. Civil Infrastruct. Eng. 18*, 5, 325–338.

GOODWIN, L. AND MAHER, S. 2000. Data mining for preterm birth prediction. In *Proceedings of the ACM Symposium on Applied Computing (SAC')*. 46–51.

GUHA, S., MEYERSON, A., MISHRA, N., MOTWANI, R., AND O'CALLAGHAN, L. 2003. Clustering data streams: Theory and practice. *IEEE Trans. Knowl. Data Eng. 15*, 3, 515–528.

GURALNIK, V. AND SRIVASTAVA, J. 1999. Event detection from time series data. In *Proceedings of the 5th International Conference on Knowledge Discovery and Data Mining (SIGKDD)*. 33–42.

HAN, J. AND KAMBER, M. 2006. *Data Mining: Concepts and Techniques* 2nd Ed., Morgan Kaufmann, San Francisco, CA.

HUANG, P., FELDMANN, A., AND WILLINGER, W. 2001. Timescales and stability: A non-instrusive, wavelet-based approach to detecting network performance problems. In *Proceedings of the 1st ACM SIGCOMM Workshop on Internet Measurement*. 213–227.

HUHTALA, Y., KARKKAINEN, J., AND TOIVONEN, H. 1999. Mining for similarities in aligned time series using wavelets. In *Proceedings of the SPIE Conference on Data Mining and Knowledge Discovery: Theory, Tools, and Technology (SPIE)*. B. V. Dasarathy, Ed. 150–160.

JACOBS, C. E., FINKELSTEIN, A., AND SALESIN, D. H. 1995. Fast multiresolution image querying. In *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*. 277–286.

JENSEN, A. AND COUR-HARBO, A. L. 2001. *Ripples in Mathematics: The Discrete Wavelet Transforms*. Springer.

KAUFMAN, L. AND ROUSSEEUW, P. J. 1990. *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley, New York, NY.

KEOGH, E. J., CHAKRABARTI, K., PAZZANI, M., AND MEHROTRA, S. 2001. Dimensionality reduction for fast similarity search in large time series databases. *Knowl. Inform. Syst. 3*, 3, 263–286.

KEOGH, E. J., CHU, S., HART, D., AND PAZZANI, M. 2004a. Segmenting time series: A survey and novel approach. In *Data Mining in Time Series Database*. World Scientific Publishing, Singapore, 1–21.

KEOGH, E. J. AND KASETTY, S. 2002. On the need for time series data mining benchmarks: A survey and empirical demonstration. In *Proceedings of the 8th International Conference on Knowledge Discovery and Data Mining (SIGKDD)*. 102–111.

KEOGH, E. J., LONARDI, S., AND CHIU, B. Y.-C. 2002. Finding surprising patterns in a time series database in linear time and space. In *Proceedings of the 8th International Conference on Knowledge Discovery and Data Mining (SIGKDD)*. 550–556.

KEOGH, E. J., LONARDI, S., AND RATANAMAHATANA, C. A. 2004b. Towards parameter-free data mining. In *Proceedings of the 10th International Conference on Knowledge Discovery and Data Mining (SIGKDD)*. 206–215.

KLIMENKO, S., MITSELMAKHER, G., AND SAZONOV, A. 2002. A cross-correlation technique in wavelet domain for detection of stochastic gravitational waves. Tech. rep. gr-qc/0208007. University of Florida, Gainesville, FL, 1–15.

KOBAYASHI, K. AND TORIOKA, T. 1994. A wavelet neural network for function approximation and network optimization. In *Proceedings of the Conference on Artificial Neural Networks in Engineering (ANNIE)*. AMSE Press, New York, NY, 505–510.

KORN, F., JAGADISH, H. V., AND FALOUTSOS, C. 1997. Efficiently supporting ad hoc queries in large datasets of time sequences. In *Proceedings of the International Conference on Management of Data (SIGMOD)*. J. Peckham, Ed. 289–300.

LAINE, A. AND FAN, J. 1993. Texture classification by wavelet packet signatures. *IEEE Trans. Patt. Anal. Mach. Intell. 15*, 11, 1186–1191.

LALITHA, E. M. 2004. Real-time multi-resolution decomposition of degrading fault signals using entropy measure. In *Proceedings of World Academy of Science, Engineering, and Technology Conference (PWASET)*. 52–55.

LAMBROU, T., KUDUMAKIS, P., SPELLER, R., SANDLER, M., AND LINNEY, A. 1998. Classification of audio signals using statistical features on time and wavelet tranform domains. In *Proceedings of the IEEE International Conference on Acoustic, Speech, and Signal Processing (ICASSP)*.

LANE, T. AND BRODLEY, C. E. 1999. Temporal sequence learning and data reduction for anomaly detection. *ACM Trans. Inform. Syst. Secu. 2*, 3, 295–331.

LEE, S.-L., CHUN, S.-J., KIM, D.-H., LEE, J.-H., AND CHUNG, C.-W. 2000. Similarity search for multidimensional data sequences. In *Proceedings of the IEEE 16th International Conference on Data Engineering (ICDE)*. 599–608.

LEE, W. AND STOLFO, S. J. 1998. Data mining approaches for intrusion detection. In *Proceedings of the 7th USENIX Security Symposium (Security)*. 79–94.

LI, C.-S., YU, P. S., AND CASTELLI, V. 1998. MALM: A framework for mining sequence database at multiple abstraction levels. In *Proceedings of the 7th International Conference on Information and Knowledge Management (CIKM)*. 267–272.

LI, G. AND KHOKHAR, A. A. 2000. Content-based indexing and retrieval of audio data using wavelets. In *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME)*. 885–888.

LI, S.-T., CHOU, S.-W., AND PAN, J.-J. 2000a. Multi-resolution spatio-temporal data mining for the study of air pollutant regionalization. In *Proceedings of the 33rd Hawaii International Conference on System Sciences*.

LI, T., LI, Q., ZHU, S., AND OGIHARA, M. 2003. A survey on wavelet applications in data mining. *ACM SIGKDD Explor. Newsl. 4*, 2, 49–68.

LI, X., DONG, S., AND YUAN, Z. 1999. Discrete wavelet transform for tool breakage monitoring. *Int. J. Mach. Tools Manufact. 39*, 1935–1944.

LI, X., TSO, S. K., AND WANG, J. 2000b. Real-time tool condition monitoring using wavelet transforms and fuzzy techniques. *IEEE Trans. Syst. Man Cybern. Part C: Appl. Rev. 30*, 3, 352–357.

LIABOTIS, I., THEODOULIDIS, B., AND SARAEE, M. 2006. Improving similarity search in time series using wavelets. *Int. J. Data Warehou. Min. 2*, 2, 55–81.

LIN, J., VLACHOS, M., KEOGH, E., AND GUNOPULOS, D. 2004. Iterative incremental clustering of time series. In *Proceedings of the 9th International Conference on Extending Database Technology (EDBT)*. 106–122.

LOTRIC, U. 2004. Wavelet based denoising integrated into multilayered perceptron. *Neurocomput. 62*, 179–196.

LUO, J., BRIDGES, S. M., AND VAUGHN, R. B. 2001. Fuzzy frequent episodes for real-time intrusion detection. In *Proceedings of the 10th IEEE International Conference on Fuzzy Systems.* 368–371.

MA, J. AND PERKINS, S. 2003. Online novelty detection on temporal sequences. In *Proceedings of the 9th International Conference on Knowledge Discovery and Data Mining (SIGKDD)*. 613–618.

MA, S. AND JI, C. 1999a. Modeling heterogeneous network traffic in wavelet domain: Part I—temporal correlation. Tech. rep. 99-03. CNAS Lab, Beijing, China, 1–32.

MA, S. AND JI, C. 1999b. *Modeling Heterogeneous Network Traffic in Wavelet Domain: Part II—non-Gaussian Traffic*. Tech. rep. 99-04. CNAS Lab, Beijing, China, 1–30.

MA, S. AND JI, C. 2001. Modeling heterogeneous network traffic in wavelet domain. *IEEE Trans. Netw. 9*, 5, 634–649.

MAGNAGHI, A., HAMADA, T., AND KATSUYAMA, T. 2004. A wavelet-based framework for proactive detection of network misconfigurations. In *Proceedings of the ACM SIGCOMM Workshop on Network Troubleshooting*. 253–258.

MALLAT, S. G. 1989. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Patt. Analy. Mach. Intell. 11*, 7, 674–693.

MALLAT, S. G. AND HWANG, W. L. 1992. Singularity detection and processing with wavelets. *IEEE Trans. Inform. Theor. 38*, 2, 617–643.

MAN, P. W. P. AND WONG, M. H. 2001. Efficient and robust feature extraction and pattern matching of time series by a lattice structure. In *Proceedings of the 10th Conference on Information and Knowledge Management (CIKM)*. 271–278.

MANDAL, M. K., ABOULNASR, T., AND PANCHANATHAN, S. 1999. Fast wavelet histogram techniques for image indexing. *Comput. Vis. Image Understand. 75*, 1-2, 99–110.

MARSLAND, S. 2001. On-line novelty detection through self-organisation with application to inspection robotics. Department of Computer Science, University of Manchester, Manchester, U.K.

MISITI, M., MISITI, Y., OPPENHEIM, G., AND POGGI, J.-M. 2005. *Wavelet Toolbox User's Guide*. The MathWorks, Inc., Natick, MA, 29–59.

MOJSILOVIC, A., POPOVIC, M., NESKOVIC, A. N., AND POPOVIC, A. D. 1997. Wavelet image extension for analysis and classification of infarcted myocardial tissue. *IEEE Trans. Biomed. Eng. 44*, 9, 856–866.

MURTAGH, F., STARCK, J.-L., AND RENAUD, O. 2004. On neuro-wavelet modeling. *Decis. Supp. Syst. J. 37*, 475–484.

NATSEV, A., RASTOGI, R., AND SHIM, K. 1999. WALRUS: A similarity retrieval algorithm for image databases. In *Proceedings of the International Conference on Management of Data (SIGMOD)*. 395–406.

NILSSON, M., FUNK, P., AND XIONG, N. 2005. Clinical decision support by time series classification using wavelets. In *Proceedings of the International Conference on Enterprise Information Systems (ICEIS)*. 24–28, 2005, 169–175.

OGDEN, T. 1997. On preconditioning the data for the wavelet transform when the sample size is not a power of two. *Commun. Statis. Part B—Sim. Comp. 26*, 467–486.

ORFANIDIS, S. J. 1996. *Introduction to Signal Processing*. Prentice Hall, Englewood Cliffs, N.J.

PERCIVAL, D. B. AND WALDEN, A. T. 2000. *Wavelet Methods for Time Series Analysis*. Cambridge University Press, Cambridge, U.K.

PETRIDIS, V., KEHAGIAS, A., PETROTH, L., BAKIRTZIS, A., MASLARIS, N., KIARTZIS, S., AND PANAGIOTOU, H. 2001. A Bayesian multiple models combination method for time series prediction. *J. Intell. Robot. Syst. 31*. 69–89.

POPIVANOV, I. AND MILLER, R. J. 2002. Similarity search over time-series data using wavelets. In *Proceedings of the 18th International Conference on Data Engineering (ICDE)*.

RENAUD, O., STARCK, J.-L., AND MURTAGH, F. 2003. Prediction based on a multiscale decomposition. *Int. J. Wavelets, Multiresolution Inform. Process. 1*, 2, 217–232.

RENAUD, O., STARCK, J.-L., AND MURTAGH, F. 2005. Wavelet-based combined signal filtering and prediction. *IEEE Trans. Syst. Man, Cybern. Part B, Cybern. 35*, 6, 1241–1251.

RIEDI, R. H., CROUSE, M. S., RIBEIRO, V. J., AND BARANIUK, R. G. 1999. A multifractal wavelet model with application to network traffic. *IEEE Trans. Informa. Theor. 45*, 4, 992–1018.

RIZZI, S. AND SARTONI, F. 1994. Medical decision support in clinical record management systems. In *Proceedings of the International Conference on Expert Systems for Development*. 267–272.

RODDICK, J. F. AND SPILIOPOULOU, M. 1999. A bibliography of temporal, spatial and spatio-temporal data mining research. *SIGKDD Explor. 1*, 1, 34–38.

SAEED, M. AND MARK, R. G. 2001. Efficient hemodynamic event detection utilizing relational databases and wavelet analysis. In *Proceedings of the Conference on Computers in Cardiology*. 153–156.

SARMA, J. 2006. *Clustercubes: Time Series Weather Prediction Using Geographic EM Clustering*. Computer Science Department, Columbia University, New York, NY, 1–5.

SCHEUNDERS, P., LIVENS, S., WOUWER, G. V. D., VAUTROT, P., AND DYCK, D. V. 1998. Wavelet-based texture analysis. *Int. J. Comput. Sci. Inform. Manage. 1*, 2, 22–34.

SHAHABI, C., CHUNG, S., AND SAFAR, M. 2001. A wavelet-based approach to improve the efficiency of multi-level surprise mining. In *Proceedings of the PAKDD International Workshop on Mining Spatial and Temporal Data*.

SHAHABI, C., TIAN, X., AND ZHAO, W. 2000. TSA-tree: A wavelet-based approach to improve the efficiency of multi-level surprise and trend queries on time-series data. In *Proceedings of the 12th International Conference on Scientific and Statistical Database Management (SSDBM)*. 55–68.

SHASHA, D. AND ZHU, Y. 2004. *High Performance Discovery In Time Series*. Springer.

SHEIKHOLESLAMI, G., CHATTERJEE, S., AND ZHANG, A. 1998. WaveCluster: A multi-resolution clustering approach for very large spatial databases. In *Proceedings of the 24th International Conference Very Large Data Bases (VLDB)*. 428–439.

SHEIKHOLESLAMI, G., CHATTERJEE, S., AND ZHANG, A. 2000. WaveCluster: a wavelet-based clustering approach for spatial data in very large databases. *VLDB J. 8*, 3-4, 289–304.

SHEIKHOLESLAMI, G., ZHANG, A., AND BIAN, L. 1999. A multi-resolution content-based retrieval approach for geographic images. *GeoInformatica Int. J. Advanc. Comput. Sci. Geograph. Inform. Syst. 3*, 2, 109–139.

SHMUELI, G. 2004. Detecting bio-terrorist attacks by monitoring multiple streams of data. In *Proceedings of the Symposium on Machine Learning for Anomaly Detection*.

SILVER, R. B. AND GINSBURG, C. M. 1984. Early prediction of the need for hospitalization in children with acute asthma. *Clin. Ped. 23*, 2, 81–84.

SOLTANI, S., BOICHU, D., SIMARD, P., AND CANU, S. 2000. The long-term memory prediction by multiscale decomposition. *Signal Process. 80*, 10, 2195–2205.

STRUZIK, Z. R. AND SIEBES, A. 1999a. The haar wavelet transform in the time series similarity paradigm. In *Proceedings of the 3rd European Conference in Principles of Data Mining and Knowledge Discovery*. 12–22.

STRUZIK, Z. R. AND SIEBES, A. 1999b. Measuring time series' similarity through large singular features revealed with wavelet transformation. In *Proceedings of the 10th International Workshop on Database and Expert Systems Applications*. 162–166.

STRUZIK, Z. R. AND SIEBES, A. P. J. M. 2000. Outlier detection and localisation with wavelet based multifractal formalism. INS-R0008. Centrum voor Wiskunde en Informatica, Amsterdarm, The Netherlands, 1–18.

SUBASI, A. 2005. Epileptic seizure detection using dynamic wavelet network. *Exp. Syst. Appl. 29*, 343–355.

SUBRAMANYA, S. R. AND YOUSSEF, A. 1998. Wavelet-based indexing of audio data in audio/multimedia databases. In *Proceedings of the International Workshop on Multi-Media Database Management Systems (MMDBMS)*. 46–53.

TZANETAKIS, G. AND COOK, P. 2002. Musical genre classification of audio signals. *IEEE Trans. Speech Audio Process. 10*, 5, 293–302.

TZANETAKIS, G., ESSL, G., AND COOK, P. 2001. Automatic musical genre classification of audio signals. In *Proceedings of the International Symposium on Music Information Retrieval (ISMIR)*. 205–210.

WANG, J. Z., WIEDERHOLD, G., AND FIRSCHEIN, O. 1997a. System for screening objectionable images using Daubechies' wavelets and color histograms. In *Proceedings of the 6th International Workshop on Interactive Distributed Multimedia Systems and Telecommunication Services (IDMS)*. M. Diaz, P. Owezarski and P. Sénac, Eds. 20–30.

WANG, J. Z., WIEDERHOLD, G., FIRSCHEIN, O., AND WEI, S. X. 1997b. Content-based image indexing and searching using Daubechies' wavelets. *Int. J. Dig. Lib. 1*, 4, 311–328.

WANG, J. Z., WIEDERHOLD, G., FIRSCHEIN, O., AND WEI, S. X. 1997c. Wavelet-based image indexing techniques with partial sketch retrieval capability. In *Proceedings of the 4th Forum on Research and Technology Advances in Digital Libraries (ADL)*. 13–24.

WANG, Y. 1995. Jump and sharp cusp detection by wavelets. *Biometrika 82*, 2, 385–397.

WANG, Y.-P., WANG, Y., AND SPENCER, P. 2006. A differential wavelet-based noise reduction approach to improve clustering of hyperspectral Raman imaging data. In *Proceedings of the 3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro.* 988–991.

WEI, L., KUMAR, N., LOLLA, V., KEOGH, E., LONARDI, S., RATANAMAHATANA, C. A., AND HERLE, H. V. 2005a. A practical tool for visualizing and data mining medical time series. In *Proceedings of the 18th IEEE Symposium on Computer-Based Medical Systems (CBMS)*. 341–346.

WEI, L., KUMAR, N., LOLLA, V. N., KEOGH, E., LONARDI, S., AND RATANAMAHATANA, C. A. 2005b. Assumption-free anomaly detection in time series. In *Proceedings of the 17th International Scientific and Statistical Database Management Conference (SSDBM)*.

WEIGEND, A. S. AND GERSHENFELD, N. A. 1994. *Time Series Prediction: Forecasting the Future and Understanding the Past*. Addison-Wesley Publishing Company, Reading, MA.

WONG, W.-K. 2004. *Data Mining for Early Disease Outbreak Detection*. School of Computer Science, Carnegie Mellon University, Pittsburgh, PA.

WU, Y.-L., AGRAWAL, D., AND ABBADI, A. E. 2000. A comparison of DFT and DWT based similarity search in time-series databases. In *Proceedings of the 9th International Conference on Information and Knowledge Management (CIKM)*. 488–495.

YAO, Y., LI, X., AND YUAN, Z. 1999. Tool wear detection with fuzzy classification and wavelet fuzzy neural network. *Int. J. Mach. Tools Manufact. 39*, 1525–1538.

YOON, H., YANG, K., AND SHAHABI, C. 2005. Feature subset selection and feature ranking for multivariate time series. *IEEE Trans. Know. Data Eng. 17*, 9, 1186–1198.

ZEIRA, G., MAIMON, O., LAST, M., AND ROKACH, L. 2004. Change detection in classification models induced from time series data. In *Data Mining in Time Series Databases*. World Scientific Publishing, Singapore, 101–125.

Received September 2008; revised January 2009; accepted April 2009

View publication stats