

Signal Processing Techniques in Genomic Engineering

XIN-YU ZHANG, FEI CHEN, STUDENT MEMBER, IEEE, YUAN-TING ZHANG, SENIOR MEMBER, IEEE, SHANNON C. AGNER, STUDENT MEMBER, IEEE, METIN AKAY, SENIOR MEMBER, IEEE, ZU-HONG LU, MARY MIU YEE WAYE, AND STEPHEN KWOK-WING TSUI

Invited Paper

Now that the human genome has been sequenced, the measurement, processing, and analysis of specific genomic information in real time are gaining considerable interest because of their importance to better the understanding of the inherent genomic function, the early diagnosis of disease, and the discovery of new drugs. Traditional methods to process and analyze deoxyribonucleic acid (DNA) or ribonucleic acid data, based on the statistical or Fourier theories, are not robust enough and are time-consuming, and thus not well suited for future routine and rapid medical applications, particularly for emergency cases. In this paper, we present an overview of some recent applications of signal processing techniques for DNA structure prediction, detection, feature extraction, and classification of differentially expressed genes. Our emphasis is placed on the application of wavelet transform in DNA sequence analysis and on cellular neural networks in microarray image analysis, which can have a potentially large effect on the real-time realization of DNA analysis. Finally, some interesting areas for possible future research are summarized, which include a biomodel-based signal processing technique for genomic feature extraction and hybrid multidimensional approaches to process the dynamic genomic information in real time.

Keywords—Biomodel-based method, bionic wavelet transform (BWT), cellular neural network (CNN), deoxyribonucleic acid (DNA) microarray, image processing, multidimensional analysis, wavelet transform (WT).

Manuscript received April 30, 2002; revised September 8, 2002. This work was supported by the Hong Kong Innovation and Technology Fund (ITS/114/01), Standard Telecommunications Ltd., and IDT Technology Ltd.

X.-Y. Zhang, F. Chen, and Y.-T. Zhang are with the Joint Research Center for Biomedical Engineering, Chinese University of Hong Kong, Hong Kong, China (e-mail: ytzhang@ee.cuhk.edu.hk).

S. C. Agner and M. Akay are with the Thayer School of Engineering, Dartmouth College, Hanover, NH 03755 USA.

Z.-H. Lu is with the National Laboratory of Molecular and Biomolecular Electronics, Southeast University, Nanjing, China.

M. M. Y. Waye and S. K.-W. Tsui are with the Department of Biochemistry, Chinese University of Hong Kong, Hong Kong, China.

Digital Object Identifier 10.1109/JPROC.2002.805308

I. INTRODUCTION

Genomic engineering is a quickly evolving interdisciplinary field that blends bioscience, medicine, and engineering. Genomic detection and analysis, as an important branch of genomic engineering, is of crucial significance in understanding the information contained in the biological genome. A major challenge for genomic research for the next few years is to elucidate the relations among sequence, structure, and function of genes. Two technologies play predominant roles in extraction and interpretation of genomic information: deoxyribonucleic acid (DNA) sequence analysis and DNA microarray analysis. DNA sequence analysis technology has been developing for decades to unravel the structure-related information, i.e., to reveal some hidden structure, to distinguish coding from noncoding regions in DNA sequence, and to explore structural similarity among DNA sequences. DNA microarray technology is one of the major breakthroughs in genomic research owing to its parallel processing feature. Microarray analysis helps in the monitoring of gene expression for tens of thousands of genes simultaneously. It has found numerous applications in the medical and biological field, such as gene discovery, disease diagnosis, drug discovery, and toxicological researches [1].

An important aspect of genomic analysis is to process DNA or ribonucleic acid (RNA) and other proteins for the detection, extraction, and classification of genomic information. Genomic information is expressed digitally in nature. For example, DNA sequences are encoded by four nitrogenated bases: adenine, thymine, cytosine, and guanine. Similarly, protein molecules are encoded by 20 types of amino acids. Both DNA and protein molecules can be mathematically represented by character strings. The character string can be properly mapped into one or more numerical sequences, then signal processing techniques provide a set of novel and useful tools for solving highly relevant problems. In a DNA microarray, the gene level is indirectly reflected in fluorescence image and characterized

by intensity signal on each spot. By means of image signal analysis, gene expression information can be extracted, compared, and classified within a microarray or among microarrays. From this point of view, the multidimensional processing techniques, if properly applied, are able to exploit the full potential of DNA sequencing and microarray technologies, and are sure to yield significant effects on the analysis of genomic data.

A number of signal processing techniques have been successfully applied to DNA sequence analysis [2]–[4]. However, the timely processing and interpretation of the vast genomic information from sequences still remains a big challenge, and some important technical problems remain unsolved. For example, immediate attention is needed to adopt proper signal processing techniques to perform accurate and automatic analysis of sequences. Thus, with the development and maturity of biosignal processing techniques, it is necessary to introduce some novel and advanced signal processing techniques into genomic analysis to facilitate research in this promising field.

Image signal processing is a critical aspect of microarray analysis. Although the basic goal—to extract the intensities signal on each spot for further analysis—is straightforward, the high density of spots as well as the variation and noise on a microarray image make it a complex process. One of the major problems that microarray signal processing faces is to develop a set of algorithms that would be more accurate and robust to the variation and noise and make possible automatic processing as well. On the other hand, the increasingly higher output of microarrays for biological and medical uses is making the processing efficiency a prominent issue in microarray signal processing. Furthermore, microarray analysis by a traditional digital computer, which sequentially processes an image pixel by pixel, not only is very time-consuming, but also destroys the parallel nature of the microarray technique itself. In this sense, finding out a parallel processing technique for real-time analysis means a breakthrough in microarray techniques. How to solve these problems successfully is one of the many interesting areas of current research in microarray signal processing.

The inference of genetic interactions for measured expression data is one of the most challenging tasks of modern functional genomics. Microarray data are created by observing gene expression values (or “activation” values) over a number of timesteps. They are often obtained while subjecting the cell to some stimulus. The goal of analyzing such data is to determine the gene regulatory network which identifies which genes have a role in exciting or inhibiting the activity of other genes in the next timestep. Thus, it would be useful to construct a three-dimensional (3-D) or four-dimensional (4-D) model to demonstrate such dynamic interaction among many mutually interacting genes.

In this article, we present an overview of the applications of recent signal processing techniques on genomic signal analysis, with the emphasis on DNA sequence analysis and DNA microarray analysis. The methods examined mainly focus on wavelet transform (WT) and cellular neural network (CNN) [5], [6] owing to their remarkable advantages, i.e., multiresolution analysis and parallel processing in real time. In addition, this article also discusses some important issues

in genomic signal processing, such as the genomic signal modeling, real-time analysis, and noise reduction, suggesting some interesting areas for possible future research.

The article is organized as follows. In Section II, we review the recent applications of existing signal processing techniques, especially WT in DNA sequence analysis for sequence structure prediction, sequence comparison, and its classification. In Section III, we examine the existing signal processing techniques for DNA microarray analysis with the emphasis on the CNN for real-time DNA analysis. Then we discuss some relevant signal processing techniques, summarizing future directions in genomic signal processing research in Section IV.

II. SIGNAL PROCESSING TECHNIQUES IN DNA SEQUENCE ANALYSIS

The DNA sequence contains the instructions that control nearly everything about how an organism lives, such as its development, metabolism, and sensitivity to infection [3]. Its analysis is an important research project in genomic signal processing. With the exponential generation of complete DNA sequences, it is particularly urgent for us to decode these inherent sequence features. Many studies have been carried out to extract the characteristic segments, to reveal some hidden structures, to distinguish coding from non-coding regions in DNA sequences, and to explore structural similarity among DNA sequences. Signal processing will play an important role in reaching this goal, and indeed many computational techniques have already been applied, including the artificial neural network (ANN) [7]–[11], nonlinear model [12], spectrogram [2], and statistical techniques [13]. In this section, the applications of WT in DNA sequence analysis will be reviewed below separately according to their different analysis tasks.

A. Sequence Structure Prediction

Accurate prediction and detection of the DNA regions or their underlying structural patterns are constant difficulties for researchers. Traditional structure detection methods were mainly based on the average of DNA base contents within a fixed window. Therefore, the location accuracy depended on the chosen window length. The multiresolution analysis feature of WT is excellent in resolving this problem, allowing efficient extraction of basic components at different scales.

Lio and Vannucci [14] applied discrete wavelet transform (DWT) to find pathogenicity islands and gene mutation events in genome data. They used DWT to smooth $G + C$ profiles to locate characteristic patterns in genome sequences, and a wavelet scalogram was obtained to compare the sequence profile among genomes and to separate the different components within a profile. Fig. 1(a) and (b) present a $G + C$ profile of the *Deinococcus radiodurans* chromosome I sequence and its scalogram used for profile comparison. Fig. 1(c) and (d) show the scalograms for two strains of *Helicobacter pylori*. The low- and high-frequency components generated from the WT-based scalogram correspond to large (cluster of genes) and small (single-gene) regions, respectively.

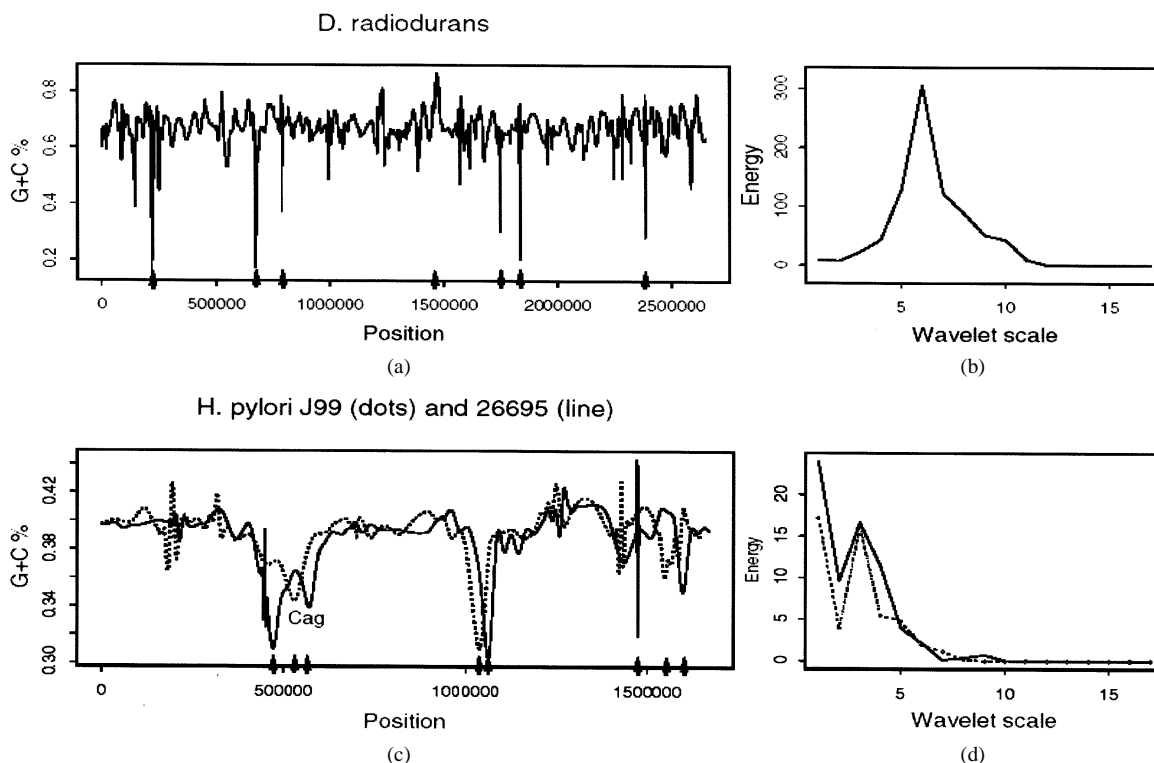


Fig. 1. (a) The $G + C$ profiles of *D. radiodurans* chromosome I, (b) with its wavelet scalogram, (c) two *H. pylori* sequences, and (d) their relative scalograms [14].

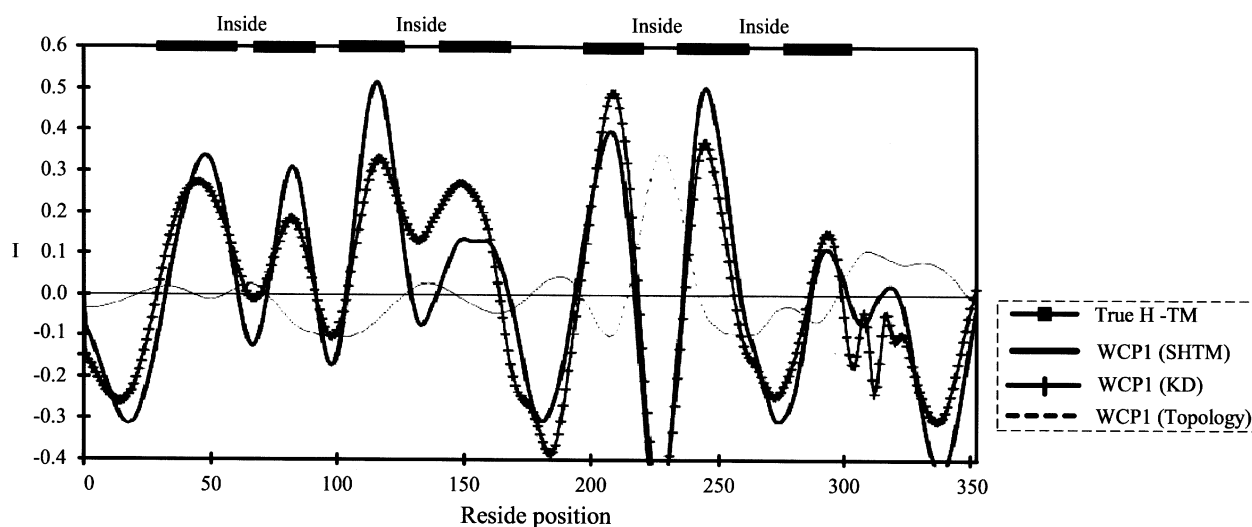


Fig. 2. Human CKR5 profiles generated by WCP and three scale presentations. The true HTM locations are shown in the topmost line [15].

Lio *et al.* [15] used a change-point based wavelet thresholding method (WCP) to predict transmembrane helix (HTM) locations and topology of HTM segments in the primary amino acid sequences. WT was applied to decompose the propensity profile, which was generated according to the frequency of residues in HTM sequences. With the wavelet coefficients, a data-dependent threshold was then used to choose the coefficients representing abrupt changes in the profile. Fig. 2 shows the WCP analysis of human CKR5 profiles using three scaling methods with different definitions. Their reported prediction results

were comparable to other methods, such as ANN and the hidden Markov model. However, the WT-based method was nonparametric in that it did not require any special model for amino acid residues, and it used change-point statistics to predict the ends of transmembrane segments. Moreover, its computational task is simple. Similarly, Pattni *et al.* [16] used continuous wavelet transform (CWT) to predict the α -helix content from the secondary structure of protein using the information from its hydrophobicity profile and the amino acid composition. Hirakawa *et al.* [17] applied wavelet analysis to predict the hydrophobic

core of proteins. From their results, the prediction accuracy was about 70%–80%; however, these prediction schemes did not require sequence alignment, which is a significant advantage in genomic signal processing [3].

Dasgupta *et al.* [18] designed models to identify the gene locations in human DNA, including the Markov model, the hidden Markov model, and a wavelet-based hidden Markov tree (HMT). In HMT processing, Dasgupta *et al.* designed adaptive wavelets matching individual CpG islands in a DNA sequence to optimize the location identification. Morozov *et al.* [19] introduced a model-based method combined with WT to depict the replacement rate variation in genes and proteins, in which the profile of relative replacement rates along the length of a sequence was defined as a function of the site number. Besides better performance in fitting the data compared with other models (e.g., the discrete gamma model), their model also provided an additional useful way for determining regions in genes and proteins that evolved significantly faster or slower than the average sequence.

B. Sequence Comparison and Classification

To understand the relationship of DNA structure and function, it is necessary to find the similarity of DNA sequences, especially for newly identified ones. For this aim, wavelet analysis will provide a useful visual description of the inherent structure underlying DNA sequences.

In [20], Trad *et al.* used wavelet analysis to extract characteristic bands from protein sequences. Their sequence-scale analysis with WT gave a multiresolution similarity comparison between protein sequences. This “similarity” expanded the traditional sequence similarity concept, which took into account only the local pairwise amino acid and disregarded the information contained in coarser spatial resolution. Also, this WT-based method did not require the complex sequence alignment processing for sequences. Therefore, proteins with different sequence lengths could be compared easily.

Sequence classification is also a major problem in DNA signal analysis. Zhao *et al.* [21] have used the wavelet packet (WP) technique for DNA sequence classification, i.e., to classify exons and introns. After obtaining the energy distribution from WP coefficients, the energy map was used as a criterion for sequence classification.

C. Exploration of the Relationship Between Sequence Structure and Function

It has also been widely agreed that a major challenge for DNA sequence analysis during the next few years will be to help elucidate the relationship between sequence structure and function of genomic sequences [22]. Scientists are exploring the structural features within the sequences. Before the wavelet method was applied, conventional Fourier analysis had been used to elucidate the sequence structure information. However, the Fourier method could discover only “global” periodicities, and it could not extract hidden localized periodicities, which might provide hints about underlying construction rules [23].

Dodin *et al.* [24] constructed a correlation function to compare each DNA base with its various neighbors. After further Fourier or WT processing applied to the correlation function,

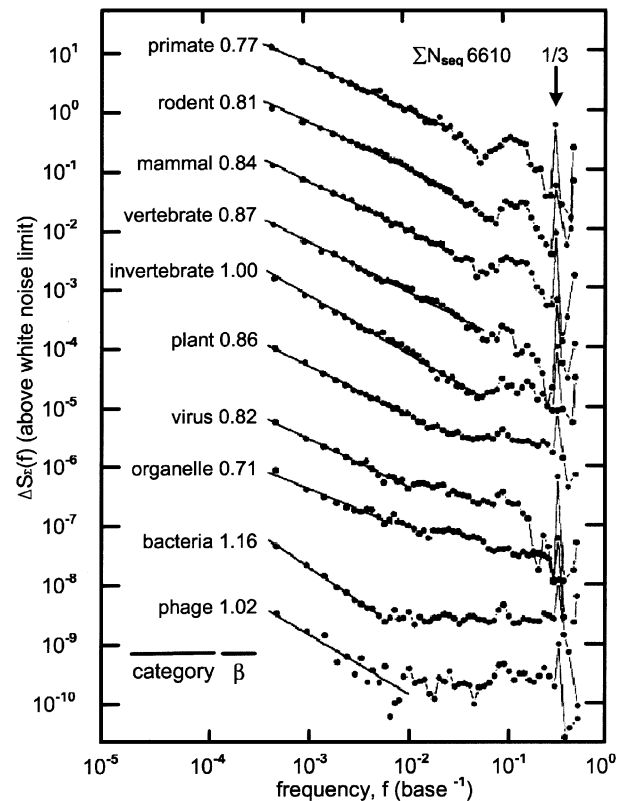


Fig. 3. Spectral density measurements for several species [25].

their results readily showed some regular features in DNA sequences. Tsonis *et al.* [23] also used WT to search the DNA sequence construction rules. The salient spots in their final two-dimensional (2-D) analysis results exhibited significant features in the DNA sequence. Their results demonstrated that while the noncoding sequences showed spectra similar to those from random sequences, coding sequences revealed specific periodicities of variable length and a common periodicity of three.

Similarly, Voss [25] applied the method in quantifying symbolic sequence correlation to analyze DNA sequences. The spectral density measurements of different base positions demonstrated the ubiquity of low-frequency $1/f^\beta$ noise, long-range fractal correlation, and prominent short-range periodicities. Voss’ results in several categories of DNA sequences also showed systematic changes in spectral exponent β . This result provides a new technique for quantifying evolutionary changes in the information content of DNA. Fig. 3 exhibits spectra approximation to $1/f^\beta$ for some categories. Changes in β are clearly seen within the categories.

Arneodo *et al.* [26] used wavelet transform modulus maxima (WTMM) to analyze the fractal scaling properties in DNA sequences. They demonstrated the existence of long-range correlation in genes containing introns and noncoding regions, and also quantified that correlation. They also found that the fluctuations in the DNA walk profiles were homogeneous with Gaussian statistics. This result reveals useful information about the role of introns and noncoding intergenic regions in the nonequilibrium dynamic process that produced DNA sequences.

Recently, it has been asserted that not only functional information can be derived from genomic data, but also information about molecular evolution and relationships between organisms [23]. Since the evolution of genetic information and the principles through which nature produced the genetic information and genes are still not well understood, wavelet analysis for DNA sequences may provide some insights for these problems.

III. IMAGE SIGNAL PROCESSING IN DNA MICROARRAY ANALYSIS

Basically, DNA microarrays [27], [28] consist of a series of DNA segments regularly arranged on some kind of support, and the expression measurement involves hybridizing the whole array with a labeled DNA or RNA sample. The principle of DNA microarray is identical with that of the traditional nucleic acid hybridization techniques. Instead of detecting and studying one gene at a time, microarrays allow thousands or tens of thousands of specific DNA or RNA sequences to be detected simultaneously on a small glass or silica slide only 1–2 cm square, and permit all of this information to appear on a single image. The uses of microarrays for gene expression profiling, genotyping, mutation detection, and gene discovery are leading to remarkable insights into the function of thousands of genes previously known only by their gene sequences. DNA microarray technology offers unprecedented advantages in postgenome research. These advantages include the ability to develop standard high-output screening methods and increased efficiency and reliability in obtaining biological molecular information. It also reduces biochemical reagent consumption as well as operation costs of previously conventional DNA sequencing techniques [29]–[35].

A DNA microarray experiment mainly consists of the following steps: probe design and microarray fabrication, sample preparation and target sequence hybridization, detection of hybridization results, and analysis for the hybridization image. The probe design actually means the selection of proper probe sequences and their arrangements on the chip and is determined by its specific goal, such as single nucleic polymorphism for a specific gene, mutation identification, or differential expression for a given group of genes. The fabrication of the DNA microarray can be done with the spotting method [36] or on-chip synthesis [37], [38]. The target genes are normally required to be amplified and labeled with fluorescence. Multicolored fluorescent techniques are often used in comparative hybridization detection. Fluorescent detection is a conventional method for the detection of hybridization result. Fig. 4 gives an example of a typical fluorescence image. The most important characteristic to draw from a fluorescence image is the assessment of the hybridization degree (e.g., whether or not, and at which quantitative level, they hybridize with a given nucleic acid species), which is proportional to the intensity of each color spot. Ratios of spot intensities in both dyes in comparative hybridization can then be used to compute the differential expression of the gene or the expressed sequence tags between the two samples. The data analysis based on fluorescence images and the gene expression

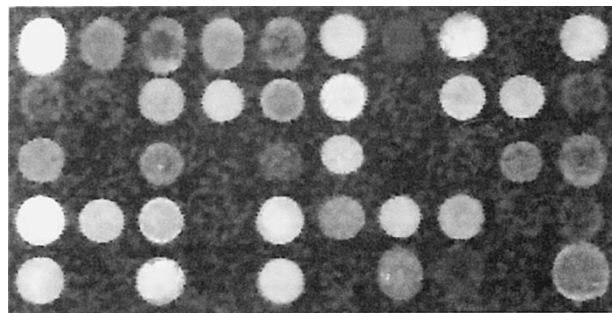


Fig. 4. A typical fluorescence image. Provided by the Biochemistry Laboratory, Chinese University of Hong Kong.

database is necessary for treatment of such a large amount of information obtained from a DNA microarray.

Since the fluorescent image obtained from microarray hybridization contains nearly all the gene expression level for detected DNA or RNA sequences, the performance of image-processing methods used for fluorescent images has a potential impact on subsequent analysis such as clustering or the identification of differentially expressed genes. Many software tools have been developed for microarray image processing. The basic goal is to transform an image of spots of varying intensities into a matrix, called a gene expression matrix, with a measure of the intensity (or, for multicolored fluorescent images, the ratio of intensities) for each spot. Although it seems to be a relatively straightforward goal, the variation, noise, and large number of pixels on a microarray image make it a complex process. The major questions these software tools face are how to reduce noise to improve the accuracy and how to realize the automation for the processing procedure. Implementing real-time processing for microarray image appears more and more urgent because of the increasing number of microarrays that must be analyzed. In this section the application of existing techniques of signal processing in microarray analysis will be examined, especially noise reduction, automation, and real-time analysis.

A. Image-Processing Methods in Microarray Analysis

For microarray image processing, the spots corresponding to genes on the microarray should be identified, their boundaries determined, and the fluorescence intensity from each spot measured and compared with background intensity. However, the automat.

The first steps in image analysis are addressing, or gridding, the image. Automation of this part permits high-output data analysis. The basic structure of a microarray image is determined by the arrayer and is therefore known. However, the automation of this process is complicated by the variation of size and position of spots, the relative placement of the adjacent grid, and the overall position of the array image. Many existing software packages for microarray image analysis require some degree of user intervention in this step. Allowing user intervention may increase the reliability and ensure accuracy of the whole quantitation process; however, this may make the process very slow. Semiautomated packages for grid generation do exist (e.g., [39]–[42]), but all require some limited degree of user intervention, either in

locating spots, setting thresholds, or making unrealistic assumptions about the regularity of data (e.g., assuming only circular spots). C. Bowman *et al.* [43] has proposed their novel operator-independent and reproducible method for the automated analysis of gene microarray images. This algorithm is based on the regular structure of the images and uses Fourier methods to extract this periodic structure as an initial approximation of spot locations. Then initial addressing is refined by an iterative method that produces accurate locations of all spots on the array.

Y. H. Yang [44] reviewed a number of existing analysis approaches, especially their use in solving the segmentation and background correction problem. He divided the existing methods for spot segmentation into four classes: fixed circle, adaptive circle, adaptive shape, and histogram. He also proposed a seeded region growing algorithm for spot segmentation with a morphological opening algorithm for background correction and made a comparison with existing analysis approaches from two sets of experiments. The comparison indicates that the choice of background adjustment method can have a large impact on the background-corrected intensity ratios that are the primary outputs of the image analysis system. In contrast, various segmentation approaches (fixed or adaptive circles or shapes) have a smaller impact. The comparison of different background correction methods indicates that background estimates based on means or medians of pixel intensities over local regions tend to be quite noisy. At the other extreme, no background adjustment seems to reduce the ability to identify differentially expressed genes. Morphological opening seems to provide a good balance—that is, less variable estimates than local background methods and more accurate estimates than raw intensities (no background correction at all)—in terms of the bias/variance trade-off.

In addition, a classifier by spot quality is employed in some software packages to automate detection of spot-finding errors and spots of poor quality for further improvement of effectiveness. Many current implementations require the user to specify explicit thresholds of various attributes, such as brightness, which separate acceptable from unacceptable spots. Choosing good thresholds manually for multiple attributes only by an extended process of trial and error is time-consuming and may not achieve the desired result. Dapple [42] implements a novel example-based classifier to decide whether candidate spots have been found correctly and are usable for further analysis. Machine learning techniques are introduced in the classifier to provide a convenient and powerful way for an investigator to specify complex concepts of spots without explicitly determining classification thresholds for image attribute values. According to their test, the automated classification matched their manual classification for more than 95% of candidate spots [42].

B. CNN Application for Real-Time DNA Microarray Analysis

Currently, with the software package mentioned above, the time spent on quantitatively processing a typical microarray fluorescent image is in the order of minutes. Some researchers think that such a speed is acceptable for laboratory use. However, it seems rather slow for a very high-output user, such as a pharmaceutical company, which might produce tens of

thousands of arrays per year. Furthermore, DNA microarray is predicted to be a normal diagnosis method in clinics in the future, much like today's blood test. A low-efficiency analysis approach is sure to impede the progress of microarray applications. Enhancing the processing speed, or ideally realizing real-time processing, is desirable. Microarray analysis by a traditional computer, which sequentially processes images pixel by pixel, not only is very time-consuming, but also destroys the parallel nature of microarray techniques itself. Fortunately, the CNN was recently introduced into microarray analysis by P. Arena *et al.*, [45]–[47], which promises to provide a breakthrough in DNA microarray parallel processing, and to obtain in real time the gene expression profile.

CNNs were introduced in 1988 by Chua and Yang [5], [6] and have been developed to overcome the massive interconnection problem of parallel distributed processing. Their key features are asynchronous parallel processing, continuous time dynamics, and local interactions among network elements. The CNN is a 2-, 3- or n -dimensional array of identical dynamical continuous systems, called cells, with only local interactions that can be programmed by the so-called template matrix [48]. Fig. 5(a) illustrates the basic 2-D CNN architecture and operating principles in a $n \times m$ CNN. As depicted in Fig. 5(c), each cell in the CNN interacts directly with the neighboring cells by means of programmable template parameters. Ideally, the dynamics of each cell is governed by the following state equation model (1), and its evolution is dependent only on the value of the time constant $\tau = RC$, which is in the order of $10^{-7} \sim 10^{-5}$ s, with an intrinsic nonlinearity between the state variable and the output of each cell [45]

$$C \frac{dv_{xij}(t)}{dt} = -\frac{1}{R_x} v_{xij}(t) + \sum_{C(k,l) \in N_r(i,j)} A(i,j;k,l) \cdot v_{ykl}(t) + \sum_{C(k,l) \in N_r(i,j)} B(i,j;k,l) \cdot v_{ukl}(t) + I_{bias}$$

$$v_{yij}(t) = \frac{1}{2} (|v_{xij}(t) + 1| - |v_{xij}(t) - 1|) \quad (1)$$

where $1 \leq i \leq m$; $1 \leq j \leq n$; A , B , and I represent the feedback, control, and bias templates, respectively, and $N_r(i, j) = \{C(k, l) | \max\{|k - i|, |l - j|\} \leq r\}$ represent the set of all neighboring cells within a radius r for the cell $C(i, j)$.

Converting the cellular processor arrays into an algorithmically programmable microprocessor architecture is the idea behind the CNN Universal Machine (CNN-UM) [49]. This is implemented by putting programmable arithmetic and logic processing units and memory units into each cell to store data. CNNs are widely applied in the field of image processing and are increasingly used in other fields because CNNs can perform parallel signal processing in real time. For more details on CNN paradigm, see [50].

For microarray applications, a set of algorithms have been developed by Arena and his colleagues using a high-level language dedicated to CNN-UM chip programming, and validated on a typical fluorescent image of DNA microarray after hybridization. To perform the microarray analysis using CNN architecture, a procedure that includes both prefiltering

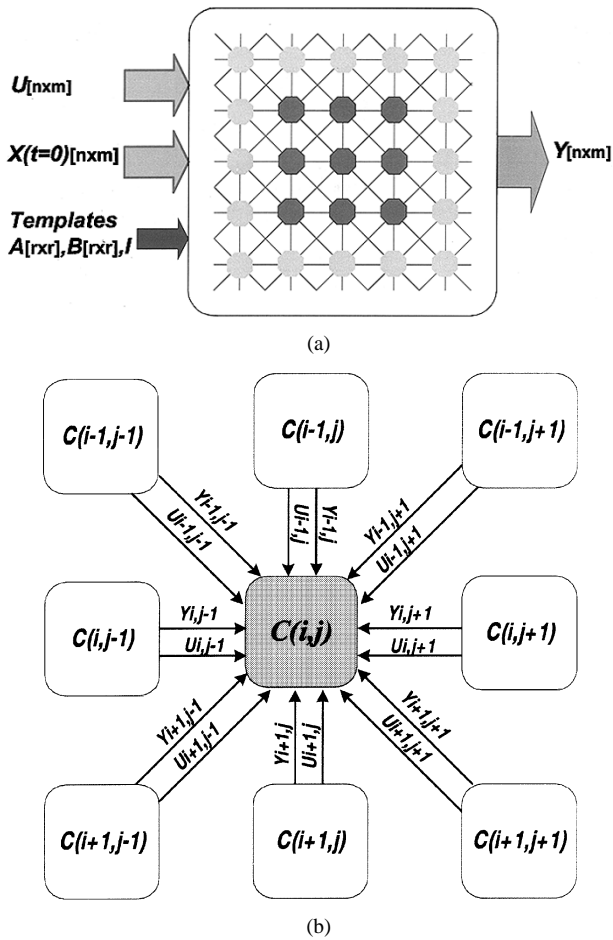


Fig. 5. (a) Block scheme of the CNN architecture and operation principle. (b) CNN local interaction between cells [45].

and the full development of the intensity analysis phase, describing each step of the algorithm in terms of single operations carried out by the CNN library of templates [51]. Fig. 6 depicts a flowchart of the fundamental steps of the image analysis procedure.

The whole algorithms are implemented on a 64 X 64 analog I/O CNN-UM chip, as shown in Fig. 7. It is reported that the on-chip time required to run the whole algorithm on a 64 X 64 color image is about 7 ms [46], which actually offers a crucial advantage with respect to currently available microarray technologies. The accuracy used for these images allowed each spot to be placed inside a square of about 16 by 16 pixels. If a large image is to be processed, the CNN-UM chip allows one to process subimages of 64 X 64 size and finally to “tile” the subresults. The input image shown in Fig. 7 thus requires six “tiling” steps to be processed; therefore, the total time consumed is about 42 ms [46]. Fig. 7 also depicts the on-chip results for the high-level red, green, and yellow spots. It is indicated that the time spent could also be further reduced, since the CNN-UM is currently a prototype and the infrastructure could be improved. However, beside the numerical results, which could be improved even further, it must be stressed that the real breakthrough lies in the new parallel processing of DNA microarrays allowed by CNNs with respect to the traditional sequential one.

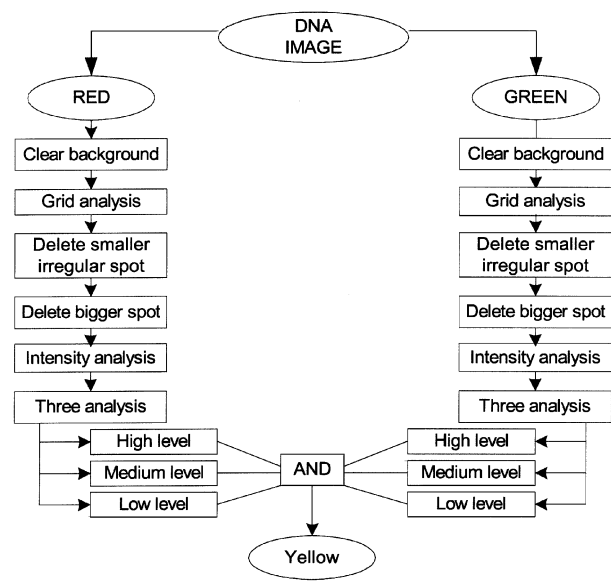


Fig. 6. Flowchart of the CNN algorithms for microarray analysis [45].

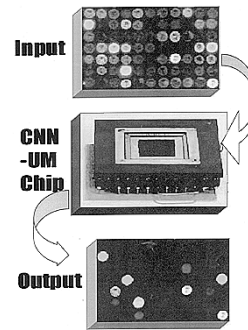


Fig. 7. CNN-UM chip for DNA microarray processing [47].

IV. DISCUSSION

A. Biomodel-Based Technique in Genomic Data Processing

With the exciting results achieved from genomic signal analysis, there also exist several open problems in this field. For instance, the prediction accuracy of characteristic genomic location is only around 70%–80%. The fixed-shape segmentation for microarray images does not match the actual cases, and an adaptive technique is needed to reliably detect the spot location and its size in microarray images [44]. Therefore, more efficient methods are being applied to solve these problems, such as singular value decomposition [52] and template technique [53], etc. In our opinion, biomodel-based approaches will play an important role in the future of genomic signal processing, since modeling methods come from the actual physiological and genomic processes, and they simulate the underlying dynamic mechanism for the signal generation. Also, these models have rich mathematical structures and form the theoretical basis for the applications. Recently, researchers have moved in this direction. Seale and Davies [54] used a stochastic model to interpret DNA microarray images, and tried to unravel their underlying physical process. Cosic [55] constructed a physicomathematical model to analyze the interaction of protein and its target.

Barral *et al.* [12] applied the nonlinear modeling method to analyze the DNA sequence. These model-based methods have provided new viewpoints in genomic information exploration and will also accelerate this process. Bionic wavelet transform (BWT) is another example of such a biomodel-based method for time-frequency analysis.

As a multiresolution signal processing method, WT can present better frequency resolution in low frequencies and better time resolution in high frequencies. However, it has its inherent limitation in 2-D analysis resolution adjustment, and behaves worse in frequency resolution, particularly in high-frequency range, in contrast to other methods, such as short-time Fourier transform (STFT), as shown in Fig. 8(a) and (b). To overcome this resolution limitation and reveal the real signal nature, BWT has been developed in our laboratory based on an active cochlea model [56]. The cochlea, one of the most crucial components in the inner ear, is responsible for frequency component separation and selectivity. Because of the active mechanism lying under the cochlea, it presents a sharp frequency-tuning feature for speech signal processing. After introducing this bionic active mechanism into WT, BWT can realize adaptive signal-dependent 2-D resolution adjustment over the time-frequency plane rather than the fixed frequency-dependent resolution in the WT case, as shown in Fig. 8(b) and (c).

Further research shows that BWT also possesses many other features, such as a more concentrated signal presentation over the time-frequency plane, and a better robustness to noise, etc. [56]. By designing an appropriate T function for different signals, BWT can efficiently reveal the signal's underlying features [56], and it should have great application potential for genomic data analysis. For example, because of its adaptive 2-D time-frequency adjustment mechanism, BWT can effectively exhibit the inherent complex structure of DNA sequence using the sharp frequency traces in its time-frequency presentation, which is useful to increase the location identification accuracy of DNA sequences. It is also helpful for the adaptive segmentation in DNA microarray image processing. Also, BWT can use fewer coefficients than WT to present the important signal details. Generally, these detailed coefficients characterize the most important signal components. Klevecz *et al.* [57] tried to use some larger wavelet coefficients to analyze the yeast cell cycle by its microarray image, and achieved useful results in uncovering the inherent dynamic architecture. Similarly, using these more concentrated coefficients of BWT, it is possible to realize efficient clustering and classification in genomic signal processing. Moreover, BWT's noise robustness feature will also reduce the noise influence in microarray image processing.

B. Usage of Multidimensional Signal Processing Technique

Until now, two independent strategies have been developed for microarray data analysis [57]. One treats the array-organized data as a one-dimensional (1-D) time-domain signal, and uses classical 1-D signal processing methods to analyze these signals. The second one treats the overall gene expression spots as a contour plot or color-mapped image, and analyzes the data in a 2-D large-scale pattern. Statistical

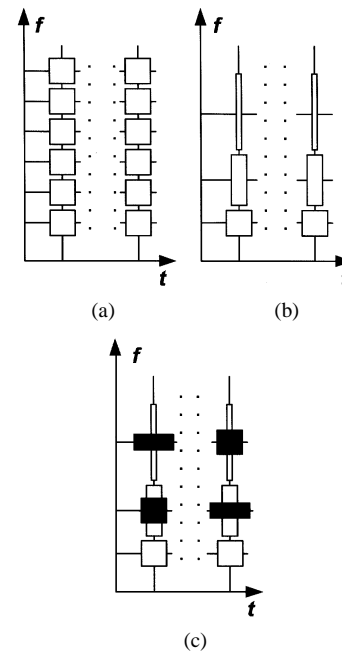


Fig. 8. Resolution comparison for (a) STFT, (b) WT, and (c) BWT. The shadowed boxes in (c) are the resolution-adjusted ones from WT.

analysis on the color distribution of the microarray image is currently a major interest and tool in microarray data processing. However, the performance of 2-D image-processing techniques will have a great impact on genomic data analysis. There still exists much scope for the application of 2-D image-processing techniques, such as 2-D WT, to improve detection accuracy. Besides, current signal processing work is just the first step in microarray information exploration, and many of the results are still qualitative. The quantitative results are of great importance for further genomic data analysis. It is also useful to extract different image patterns, and to find correlations between these image data for comparison and classification of biological processes. Meanwhile, the microarray image's own features also put forward new technical challenges for engineers to consider, such as high-speed implementation and 2-D parallel fast algorithm for real-time microarray processing.

Multidimensional (beyond 2-D) analysis will be another research trend in the genomic research field. For instance, in protein engineering research, it is of great interest to reveal the complex 3-D structure of the genomic sequence. Currently, most genomic signal processing techniques are still static. The 4-D technique will help to detect the real-time dynamic genomic structure change during future drug experiments. With the application of these multidimensional techniques, it will be possible in the future for us to reveal the underlying genomic structure and function, and their relationship in dynamic situations.

C. Noise Reduction

Noise reduction is a prominent issue in DNA microarray analysis. If a DNA microarray image is an array of spots with a precise size and position located on a uniform low background, it is a really simple problem in terms of image processing. It is the variation and noise in fluorescent images

that complicate this problem. It is important for us to analyze the source of noise and to seek effective solutions.

The major source of variation and noise on microarray images originates from microarray fabrication machines, the treatment of glass slides, and fluorescence detectors. Despite the use of precise fabrication machines, spots vary significantly in size and position owing to variations in the amount of DNA on each spot and in the location where it is spotted. Detector noise includes that from the amplification and digitization process, such as photon noise, electronic noise, laser light reflection, and background fluorescence [44]. In practice, the natural fluorescence of the glass and any nonspecifically bound DNA or dye molecules add a substantial noise floor to the image. This diffuse noise exhibits considerable variability in intensity both within and between small rectangles containing individual spots. Microarrays are also afflicted with discrete image artifacts such as highly fluorescent dust particles, unattached dye, salt deposits from evaporated solvents, and fibers or other airborne debris. Such artifacts appear in the vicinity of 10%–15% of spots at random, even after thorough cleaning of the slide, and can easily be brighter and sometimes larger than nearby useful spots [42]. Their heterogeneous brightness, shape, and size make them hard to detect and remove automatically, especially in the presence of spots that are themselves of variable size and brightness. Bright artifacts complicate spot finding because sometimes they are mistaken as spots.

Generally, noise reduction can be carried out from two aspects: first, to accurately adjust the fabrication machine and fluorescence detector as well as to normalize the experimental conditions; second, to properly design filters to extract a desired signal that is originally contaminated by noise. Filter design should start at the analysis of image features, including the morphological feature in objects of interest and the spatial frequency feature as well as local contextual features. In microarray analysis, the basic structure of the whole image is nearly invariable, and the background intensity usually varies relatively slowly across the image. In this sense, denoising wavelet analysis and transform techniques may be a better, more appealing solution for automated spot finding and background correction, since WT is a local transformation between time and frequency, and is able to effectively extract the information of interest. Recent developments with WT provide it with several advantages [58]: first, wavelet filters cover the frequency domain exactly; second, correlation between the features extracted from a distinct filter bank can be greatly reduced by an appropriately designed filter; third, adaptive pruning of the decomposition tree makes possible the reduction of the computational complexity and the length of feature vectors; and finally, fast algorithms are available to facilitate the implementation. Wavelet-based image-denoising algorithms have been widely used for different biomedical applications, and also have great potential to cope with noise reduction problems in microarray analysis.

D. Applications of CNN-Based Microprocessors in Microarray Analysis

Until now, what has been done with CNN-UM in microarray analysis is to process the fluorescence image and

measure the intensity of each spot. However, important features of CNN-UM, such as parallel computation, local interconnection between network elements, and hybrid (analog and logic) computing, have unprecedented potential in other phases of microarray analysis. It is possible that, with the future development of CNN and related technology, CNN-based microprocessors can perform nearly all the tasks in DNA microarray analysis, including image processing, comparison, classification, and analysis of gene expression.

CNN-UMs can be programmed using a dedicated high-level language, and a variety of templates are available for image processing. Using more sophisticated image-processing algorithms in CNN-UM to cope with noise problems, one should be able to increase the accuracy without destroying the real-time property, which is hard to do with conventional digital computers because of their serial computing nature.

CNN-based microprocessors can also deal with the identification, comparison, and classification within the gene expression matrix or among matrices obtained from image processing. As mentioned in Section I, to identify which genes have a role in exciting or inhibiting the activity of other genes in the next timestep, a gene regulatory network [59], [60] is often constructed, using microarray data over a number of timesteps. A number of methods have been attempted to generate the gene regulatory network for microarray data, involving Bayesian networks, clustering, statistics, visualization, weight matrices, unsupervised neural networks, supervised neural networks, and even the standard ANN [59], [61]. Since the CNN concept is partly inspired by the architecture of the neural networks, the programmability of their interconnection weights and hybrid computing provide the single-layer or multilayer CNN-UM powerful capacities to carry out gene regulatory pattern analysis in real time.

There still exists a bottleneck in the CNN's input and output with the external world [62]. In fact, it is easily understood that while the number of cells in an integrated realization grows as N^2 , the corresponding number of pins can grow only linearly. This forces a sequential input/output from the chip. If this problem cannot be effectively solved, it probably reduces interest in the CNN implementation itself, since its parallel processing capacities cannot be fully exploited, and other alternative sequential approaches would probably provide similar performance at a lower cost [63]. Many researchers have focused their efforts on the development of new generations of CNN-UM devices capable of overcoming the drawbacks of traditional ones through the incorporation of the sensory and processing circuitry in each cell, making this processing act concurrent with the acquisition of the signal. Because the hybridization result can be read out on-site, a possible solution for real-time microarray image analysis is to integrate the electrical, optical, and/or chemical sensors with the CNN-UM processors in each cell [45]. This solution can implement input and output in parallel and avoid any data conversion; thus, image processing would be greatly accelerated. The possibility of integrating the sensor directly with processing circuitry on each cell in a CNN-UM chip, which is able to perform the

signal acquisition, processing, and data analysis together, opens the way to more powerful systems for real-time microarray analysis.

In summary, some interesting areas for possible future research in genomic signal processing are as follows: 1) biomodel-based signal processing techniques for genomic feature extraction and functional classification; 2) hybrid multidimensional approaches to process the dynamic genomic information for the discovery and development of new drugs; 3) development of novel signal processing techniques for noise reduction and microarray analysis; and 4) integration of sensor and processor on each cell in a CNN chip to perform signal acquisition, processing, and data analysis for real-time DNA microarray analysis.

ACKNOWLEDGMENT

The authors would like to thank Dr. X. L. Hu and T. Ma at the Chinese University of Hong Kong for their input.

REFERENCES

[1] L. M. Shi. (2002, Jan. 7) DNA microarray (genome chips)—monitor the genome on a chip. [Online]. Available: <http://www.genchips.com>.

[2] D. Anastassiou, "Genomic signal processing," *IEEE Signal Processing Mag.*, vol. 18, pp. 8–20, July 2001.

[3] S. L. Salzberg, "Gene discovery in DNA sequences," *IEEE Intell. Syst.*, vol. 14, pp. 44–48, Nov./Dec. 1999.

[4] W. Wang and D. H. Johnson, "Computing linear transforms of symbolic signals," *IEEE Trans. Signal Processing*, vol. 50, pp. 628–634, Mar. 2002.

[5] L. O. Chua and L. Yang, "Cellular neural networks: Theory," *IEEE Trans. Circuits Syst. I*, vol. 35, pp. 1257–1272, Oct. 1988.

[6] ———, "Cellular neural networks: Application," *IEEE Trans. Circuits Syst. I*, vol. 35, pp. 1273–1290, Oct. 1988.

[7] Q. C. Ma, J. T. L. Wang, D. Shasha, and C. H. Wu, "DNA sequence classification via an expectation maximization algorithm and neural networks: A case study," *IEEE Trans. Syst., Man, Cybern. C*, vol. 31, pp. 468–475, Nov. 2001.

[8] L. M. Fu, "An expert network for DNA sequence analysis," *IEEE Intell. Syst.*, vol. 14, pp. 65–71, Jan./Feb. 1999.

[9] A. Hatzigeorgiou, N. Mache, and M. Rezczo, "Functional site prediction on the DNA sequence by artificial neural networks," in *Proc. IEEE Int. Joint Symp. Intell. Syst.*, 1996, pp. 12–17.

[10] H. Ogura, H. Agata, M. Xie, T. Odaka, and H. Furutani, "A study of learning splice sites of DNA sequence by neural networks," *Comput. Biol. Med.*, vol. 27, pp. 67–75, Jan. 1997.

[11] E. C. Ueberbacher and R. J. Mural, "Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach," *Proc. Nat. Acad. Sci.*, vol. 88, pp. 11 261–11 265, Dec. 1991.

[12] J. P. Barral, A. Hasmy, J. Jimenez, and A. Marcano, "Nonlinear modeling technique for the analysis of DNA chains," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdisp. Top.*, vol. 61, pp. 1812–1815, Feb. 2000.

[13] K. Essien, M. Akay, and M. Sekine, "Investigation of protein similarity using the Kolmogorov–Smirnov test," in *Proc. IEEE Special Top. Conf. Mol., Cell., Tissue Eng. 2002*, pp. 38–39.

[14] P. Lio and M. Vannucci, "Finding pathogenicity islands and gene transfer events in genome data," *Bioinformatics*, vol. 16, pp. 932–940, Oct. 2000.

[15] ———, "Wavelet change-point prediction of transmembrane proteins," *Bioinformatics*, vol. 16, pp. 376–382, Apr. 2000.

[16] L. Pattini, L. Riva, and S. Cerutti, "A wavelet based method to predict the alpha helix content in the secondary structure of globular proteins," in *Proc. IEEE Special Top. Conf. Mol., Cell., Tissue Eng. 2002*, pp. 142–143.

[17] H. Hirakawa, S. Muta, and S. Kuhara, "The hydrophobic cores of proteins predicted by wavelet analysis," *Bioinformatics*, vol. 15, pp. 141–148, Feb. 1999.

[18] N. Dasgupta, S. Lin, and L. Carin, "Sequential modeling for identifying gene locations in human genome," Dept. Elec. Comput. Eng., Duke Univ., Durham, NC, Tech. Rep., Dec. 2001.

[19] P. Morozov, T. Sitnikova, G. Churchill, F. J. Ayala, and A. Rzhetsky, "A new method for replacement rate variation in molecular sequences: Application of the Fourier and wavelet models to Drosophila and mammalian proteins," *Genetics*, vol. 154, pp. 381–395, Jan. 2000.

[20] C. H. Trad, Q. Fang, and I. Cosic, "Protein sequence comparison based on the wavelet transform approach," *Protein Eng.*, vol. 15, pp. 193–203, Mar. 2002.

[21] J. Zhao, X. W. Yang, J. P. Li, and Y. Y. Tang, "DNA sequences classification based on wavelet packet analysis," in *Wavelet Analysis and Its Applications, 2nd Int. Conf. WAA*, 2001, pp. 424–429.

[22] D. Dalevi and S. G. E. Andersson, "Discovering the dynamics of microbial genomes," *IEEE Eng. Med. Bio. Mag.*, vol. 20, pp. 55–60, Apr. 2001.

[23] A. A. Tsonis, P. Kumar, J. B. Elsner, and P. A. Tsonis, "Wavelet analysis of DNA sequences," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdisp. Top.*, vol. 53, pp. 1828–1834, Feb. 1996.

[24] G. Dodin, P. Vanderghyest, P. Levoir, C. Cordier, and L. Marcourt, "Fourier and wavelet transform analysis, a tool for visualizing regular patterns in DNA sequences," *J. Theor. Biol.*, vol. 206, pp. 323–326, Oct. 2000.

[25] R. F. Voss, "Evolution of long-range fractal correlations and $1/f$ noise in DNA base sequences," *Phys. Rev. Lett.*, vol. 68, pp. 3805–3808, Jun. 1992.

[26] A. Arneodo, E. Bacry, P. V. Graves, and J. F. Muzy, "Characterizing long-range correlations in DNA sequences from wavelet analysis," *Phys. Rev. Lett.*, vol. 74, pp. 3293–3296, Apr. 1995.

[27] B. Jordan, Ed., *DNA Microarray: Gene Expression Applications*. Berlin, Germany: Springer-Verlag, 2001.

[28] T. J. Aitman, "DNA microarray in medical practice: Science, medicine, and the future," *Br. Med. J.*, vol. 323, pp. 611–615, Sept. 15, 2001.

[29] S. M. Y. Lee, M. L. Y. Li, Y. C. Tse, S. C. L. Leung, M. M. Z. Lee, S. K. W. Tsui, K. P. Fung, C. Y. Lee, and M. M. Y. Waye, "Paeoniae Radix, a Chinese herbal extract, inhibit hepatoma cells growth by inducing apoptosis in a p53 independent pathway," *Life Sci.*, vol. 71, pp. 2267–2277, 2002.

[30] L. D. S. Kok, T. C. C. Au, C. H. Yu, K. K. Chan, S. S. Siu, K. O. Ng, W. H. Yiu, M. M. Ng, M. Kotaka, M. Y. Lee, W. H. Tam, P. T. W. Law, A. H. Chan, Y. M. Lau, S. M. Ngai, C. T. Liew, J. W. Y. Lau, S. K. W. Tsui, C. Y. Lee, K. P. Fung, and M. M. Y. Waye, "DNA sequence analysis of expressed sequence tags (EST's) from human liver cancer," in *Molecular Genetic Basis of Cancer*, M. Lung and W. Hsiao, Eds. Hong Kong, China: HKUST Press, 2001, pp. 141–148.

[31] D. M. Hwang, A. A. Dempsey, R.-X. Wang, M. Rezvani, J. D. Barrans, K.-S. Dai, H.-Y. Wang, H. Ma, E. Cukerman, Y.-Q. Liu, J.-R. Gu, J.-H. Zhang, S. K. W. Tsui, M. M. Y. Waye, K.-P. Fung, C. Y. Lee, and C.-C. Liew, "A genome-based resource for molecular cardiovascular medicine. Toward a compendium of cardiovascular genes," *Circulation*, vol. 96, pp. 4146–4203, 1997.

[32] C. C. Liew, D. M. Hwang, R. X. Wang, S. H. Ng, A. Dempsey, D. H. Y. Wen, H. Ma, E. Cukerman, X. G. Zhao, Y. Q. Liu, X. K. Qiu, X. M. Zhou, J. R. Gu, S. Tsui, K. P. Fung, M. M. W. Waye, and C. Y. Lee, "Construction of a human heart cDNA library and identification of cardiovascular based genes (CVBest)," *Mol. Cellular Biochem.*, vol. 172, pp. 81–87, 1997.

[33] D. M. Hwang, Y. W. Fung, R. X. Wang, C. M. Laurensen, S. H. Ng, W. Y. Lam, K. W. Tsui, K. P. Fung, M. Waye, C. Y. Lee, and C. C. Liew, "Analysis of expressed sequence tags from a fetal human heart cDNA library," *Genomics*, vol. 30, pp. 293–298, 1995.

[34] Y. Wei, Z. Lu, and C. Yuan, "Molecular electronics: The strategies and process in China," *IEEE Eng. Med. Biol. Mag.*, vol. 16, pp. 53–61, July/Aug. 1997.

[35] Y. Wei, "Molecular electronics—The future of bioelectronics," *Supramolecular Sci.*, vol. 5, pp. 723–731, 1998.

[36] V. G. Cheung, M. Morley, F. Aguilar, A. Massimi, R. Kucherlapati, and G. Childs, "Making and reading microarrays," *Nature Genetics Suppl.*, vol. 21, pp. 15–19, Jan. 1999.

- [37] R. J. Lipshutz, S. P. A. Fodor, T. R. Gingeras, and D. J. Lockhart, "High density synthetic oligonucleotide arrays," *Nature Genetics Suppl.*, vol. 21, pp. 20–24, Jan. 1999.
- [38] Y. Xia and G. Whiteside, "Soft Lithography," *Angew. Chem. Int. Ed.*, vol. 37, pp. 550–575, 1998.
- [39] M. Eisen. (1999, Nov.) *ScanAlyze User Manual*. Stanford Univ. US. [Online]. Available: <http://rana.lbl.gov/manuals/ScanAlyzeDoc.pdf>.
- [40] Image Analysis Group. (2002, Aug.) *Spot User's Manual*. CSIRO, Australia. [Online]. Available: <http://experimental.act.cmis.csiro.au/spot/spotmanual.php>.
- [41] Media Cybernetics, Inc.. Array-pro analyzer. [Online]. Available: <http://www.mediacy.com/arraypro.htm>.
- [42] J. Buhler, T. Ideker, and D. Haynor, "Dapple: Improved techniques for finding spots on DNA microarray," *Comput. Sci. Eng., Univ. Washington, Seattle, Tech. Rep. UWTR 2000-08-05*, Aug. 2000.
- [43] C. Bowman, R. Baumgartner, and S. Booth, "Automated analysis of gene-microarray images," in *Proc. IEEE Can. Conf. Elect. Comput. Eng.*, 2002, pp. 1140–1144.
- [44] Y. H. Yang, M. J. Buckley, S. Dudirot, and T. P. Speed, "Comparison of methods for image analysis on cDNA microarray data," *J. Comput. Graph. Statist.*, vol. 11, pp. 108–136, Jan. 2002.
- [45] P. Arena, L. Fortuna, and L. Occhipinti, "A CNN algorithm for real time analysis of DNA microarrays," *IEEE Trans. Circuits Syst. I*, vol. 49, pp. 335–340, Mar. 2002.
- [46] P. Arena, M. Bucolo, L. Fortuna, and L. Occhipinti, "Cellular neural networks for real-time DNA microarray analysis," *IEEE Eng. Med. Biol. Mag.*, vol. 21, pp. 17–25, Mar./Apr. 2002.
- [47] L. Fortuna, P. Arena, D. Balya, and A. Zarandy, "Cellular neural networks: A paradigm for nonlinear spatio-temporal processing," *IEEE Circuits Syst. Mag.*, vol. 1, pp. 6–21, Apr. 2001.
- [48] L. O. Chua and T. Roska, "The CNN paradigm," *IEEE Trans. Circuits Syst. I*, vol. 40, pp. 147–156, Mar. 1993.
- [49] T. Roska and L. O. Chua, "The CNN universal machine: An analogic array computer," *IEEE Trans. Circuits Syst. II*, vol. 40, pp. 163–172, Mar. 1993.
- [50] L. O. Chua, *CNN: A Paradigm for Complexity*. Singapore: World Scientific, 1998, vol. 31.
- [51] *CNN Software Library, Ver 1.1*. Budapest, Hungary: Analogical and Neural Computing Laboratory, 2000.
- [52] O. Alter, P. O. Brown, and D. Botstein, "Singular value decomposition for genome-wide expression data processing and modeling," *Proc. Nat. Acad. Sci.*, vol. 97, pp. 10 101–10 106, Aug. 2000.
- [53] T. R. Hvidsten *et al.*, "Template-based gene expression analysis," in *Proc. 4th Ann. Int. Conf. Comput. Mol. Bio.*, 2000, pp. 11–12.
- [54] D. Seale and S. W. Davies, "Stochastic model of DNA microarray," in *Proc. IEEE Special Top. Conf. Mol., Cell., Tissue Eng. 2002*, pp. 113–114.
- [55] I. Cosic, "Macromolecular bioactivity: Is it resonant interaction between macromolecules?—Theory and applications," *IEEE Trans. Biomed. Eng.*, vol. 41, pp. 1101–1114, Dec. 1994.
- [56] J. Yao and Y. T. Zhang, "Bionic wavelet transform: A new time-frequency method based on an auditory model," *IEEE Trans. Biomed. Eng.*, vol. 48, pp. 856–863, Aug. 2001.
- [57] R. R. Klevecz, "Dynamic architecture of the yeast cell cycle uncovered by wavelet decomposition of expression microarray data," *Funct. Integr. Genomics*, vol. 1, pp. 186–192, Nov. 2000.
- [58] A. Laine and J. Fan, "Frame representation for texture segmentation," *IEEE Trans. Image Processing*, vol. 5, pp. 771–780, May 1996.
- [59] E. Keedwell, A. Narayanan, and D. Savić, "Modeling gene regulatory networks using artificial neural networks," in *Proc. Int. Joint Conf. Neural Networks 2002*, pp. 183–189.
- [60] Z. Szallasi, "Genetic network analysis in light of massively parallel biological data acquisition," in *Pacific Symp. Biocomputing*, vol. 4, 1999, pp. 5–16.
- [61] S. Knudsen, *A Biologist's Guide to Analysis of DNA Microarray Data*. New York: Wiley, 2002.
- [62] G. Manganaro, P. Arena, and L. Fortuna, *Cellular Neural Networks: Chaos, Complexity and VLSI Processing*. Berlin, Germany: Springer-Verlag, 1999, pp. 19–20.
- [63] T. Roska and A. Rodriguez-Vazquez, "Review of CMOS implementation of the CNN universal machine-type visual microprocessor," in *Proc. IEEE Int. Symp. Circuits Syst. 2000 II*, May 28–31, 2000, pp. 120–123.



Xin-Yu Zhang was born in China in 1976. She received the B.E. degree and the M.E. degree in biomedical engineering in 1999 and 2002, respectively, from Xi'an Jiaotong University, Xi'an, China.

She currently serves on the research staff in the Department of Electronic Engineering, the Chinese University of Hong Kong, Hong Kong, China. Her current research interests include biomedical image analysis, signal processing, and biosensors.



Fei Chen (Student Member, IEEE) was born in Zhenjiang, China, in 1975. He received the B.S. degree in electronic science and the M.S. degree in engineering from Nanjing University, Nanjing, China, in 1998 and 2001, respectively.

He is currently pursuing the Ph.D. degree in the Department of Electronic Engineering, the Chinese University of Hong Kong, Hong Kong, China. His Ph.D. research deals with biomodeling and its application in signal processing. His current research interests also

include bioinformatics, wavelet transform, biosignal processing, and data compression.



Yuan-Ting Zhang (Senior Member, IEEE) received the Ph.D. degree from the University of New Brunswick, Fredericton, Canada, in 1990.

From 1989 to 1994, he was a Research Associate and Adjunct Assistant Professor at the University of Calgary, Calgary, Canada. He is currently a Professor and Director of the Joint Research Centre for Biomedical Engineering at the Chinese University of Hong Kong, Hong Kong, China. His work has been published in several books, more than 20 scholarly journals, and many

international conference proceedings. His research activities have focused on the development of biomodel-based signal processing techniques to improve the performance of medical devices and biosensors, particularly for telemedicine.

Dr. Zhang served as the Technical Program Chair of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS). He was the Chairman of the Biomedical Division of the Hong Kong Institution of Engineers and the Vice-President of IEEE-EMBS in 2000 and 2001, an AdCom member of IEEE-EMBS in 1999, and the Vice-President of IEEE-EMBS in 2000 and 2001. He serves currently as an Associate Editor of IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, an Editorial Board member for the Book Series of Biomedical Engineering published by Wiley and IEEE Presses, and an Associate Editor of IEEE TRANSACTIONS ON MOBILE COMPUTING.

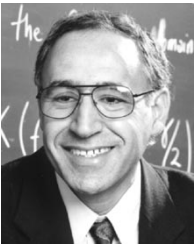


Shannon C. Agner (Student Member, IEEE) received the A.B. degree from Dartmouth College, Hanover, NH, in June 2002. She is currently working toward the masters degree in the biomedical engineering program at the Thayer School of Engineering, Dartmouth College.

Her research interests include the dynamics of respiratory patterns and the development of respiratory neural networks during maturation. She is also interested in understanding and quantification of the body motion in patients

with Parkinson's disease and the use of advanced signal processing methods for the complexity analysis of biological signals.

Ms. Agner is a student member of the IEEE Engineering in Medicine and Biology Society. As an NSF fellow, she also attended the IEEE EMBS 2nd International Summer School on Biocomplexity from System to Gene in 2002.



Metin Akay (Senior Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from the Bogazici University, Istanbul, Turkey, in 1981 and 1984, respectively, and the Ph.D. degree from Rutgers University, New Brunswick, NJ, in 1990.

He is currently Associate Professor of Engineering, Psychology and Brain Sciences, and Computer Science at Dartmouth University, Hanover, NJ. He has played a key role in promoting biomedical education in the world

by writing several prestigious books and editing the *IEEE Biomedical Engineering Book Series* (New York: Wiley/IEEE Press), sponsored by the IEEE Engineering in Medicine and Biology Society (EMBS). He is the author or coauthor of 14 books. His Neural Engineering and Informatics Lab is interested in investigating the motor functions of patients with Parkinson's disease and the effect of developmental abnormalities and maturation on the dynamics of respiration.



Zu-Hong Lu received the Ph.D. degree in bioelectronics from Southeast University, Nanjing, China, in 1988. During his doctoral period, he spent a year in University of Wales, Bangor, in the Collaboration Training Program.

He has been the Director of the National Laboratory of Molecular and Biomolecular Electronics of the Ministry of Education, Nanjing, China, since 1993, and the Director of the Institute of Science of Southeast University since 1999. He has published more than 100

research papers in international journals. He developed a microstamping technology to fabricate high-density oligonucleotide microarray chips and the related design method. He has applied for more than ten related patents. He organized three national academic meetings on the topics of molecular electronic devices, LB films, and biomedical electronics. As the co-chairperson, he organized the fourth China–Japan Bilateral Symposium on Electric-Photonic Intelligent Materials and Molecular Electronics, and the seventh International Conference on Molecular Electronics and Bio-computing. His previous research interests include dielectric and electric studies of hydrated proteins, ultrathin organic films and its applications, molecular devices, and biosensors. His current research interests include microarray technology and bioinformatics.

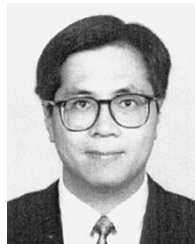
Dr. Lu has received several academic awards from the Ministry of Education. He is a General Secretary of the Society of Biomedical Electronics, Chinese Institute of Electronics, and a Member of several international societies, including the IEEE Engineering in Medicine and Biology Society.



Mary Miu Yee Waye received the B.Sc. degree (Hon.) in bacteriology and immunology from the University of Western Ontario, London, Canada, and the Ph.D. degree in medical biophysics from the University of Toronto, Toronto, Canada. From 1982 to 1985 she was a postdoctoral fellow of the National Cancer Institute of Canada at the Medical Research Council, Laboratory of Molecular Biology, Cambridge, England.

From 1986 to 1992, she was an Associate Member of the Medical Research Council Group in periodontal physiology and Assistant Professor at the University of Toronto. In 1992, she joined the Faculty of Medicine, Department of Biochemistry, Chinese University of Hong Kong, Hong Kong, China. In 2000, she became a Professor at the Chinese University of Hong Kong, where she studies the gene expression profiles of heart and liver diseases. She has identified and characterized a series of novel human genes, including the LIM domain protein family. She directs a genomic research group that provides a potential source of new markers for diseases. She has chaired the Exchange Students Committee of Chung Chi College, Chinese University of Hong Kong, since 1999.

In 1982, Dr. Waye was named the NCI King George V Silver Jubilee Cancer Fellow. She is one of the founding members of the Hong Kong Bioinformatics Centre.



Stephen Kwok-Wing Tsui received the B.Sc. degree in chemistry and the Ph.D. degree in biochemistry from the Chinese University of Hong Kong, Hong Kong, China, in 1985 and 1995, respectively. In conjunction with his collaborators at the University of Toronto, Toronto, Canada, he and his colleagues had sequenced more than 80 000 human gene fragments, and more than 2×10^7 base pairs of DNA sequence data were generated.

In 1997, he was Assistant Professor, Biochemistry Department, Chinese University of Hong Kong. He is currently Associate Professor, Biochemistry Department, Faculty of Medicine, Chinese University of Hong Kong. In 1998, he and his colleagues received a research grant (more than US\$1 200 000) from the Industry Department of the government of Hong Kong to establish a catalogue of genes in liver cancer and mental diseases. Recently, he successfully produced his own microarrays using the cDNA pool established in his laboratory. He has named and characterized more than 15 human genes, and has published more than 40 scientific papers in international refereed journals. His research interests include investigating molecular events behind human cardiovascular diseases and liver cancer, the regeneration of cardiac cells, the applications of cDNA microarrays, and bioinformatics.