Application of Wavelet Transforms to Chromosome Images

Wenzhong YAN*1, Lei BAI 1

1 Department of Computer, North China Institute of Science and Technology, Sanhe, Hebei, 065201, P.R. China

Abstract — Chromosome karyotyping refers to the classification and subsequent formatted display of the chromosomes found in a cell spread. A karyotype is required to assign each chromosome to one of 24 classes. Since karyotyping is a time-consuming procedure, computer-based classifiers have been proposed. A large number of studies have been conducted to develop computer-assisted chromosome detection and classification systems, as well as to evaluate and improve their performance. Wavelet transforms use different basis functions leading to the characterization and localization of image features in both the space and transform domains simultaneously. They offer the means of image representation in order to facilitate multiresolution image analysis. Many researchers have set out to examine the use of wavelet transforms for chromosome enhancement and classification. Multi-color fluorescence in situ hybridization (M-FISH) imaging is a recently developed molecular cytogenetic technique for the rapid visualization of genomic aberrations at the subcellular level. Some researchers have already proposed algorithms for M-FISH using wavelet transforms. This paper is focused on these proposed algorithms. The principle and the realization of these algorithms are analyzed.

Keywords - chromosome; enhancement; classification; wavelet; M-FISH.

I. INTRODUCTION

Chromosomes are genetic information carriers and their analysis is an important procedure in clinical and cancer cytogenetics studies. Chromosome karyotyping refers to the classification and subsequent formatted display of the chromosomes found in a cell spread. A karyotype is required to assign each chromosome to one of 24 classes (22 autosomes and two sex chromosomes). Fig.1 shows a sample karyotype. Since karyotyping is a lengthy process, computer-based classifiers have been proposed. Since the 1980s, automated chromosome detection and classification have attracted great research interest. A large number of studies have been conducted to develop computer-assisted chromosome detection and classification systems, as well as to evaluate and improve their performance. A computer-assisted system usually includes four processing steps: (1) image enhancement, (2) chromosome segmentation (detection) and alignment, (3) feature computation and selection and (4) chromosome classification [1, 2, 3].

Wavelet transforms use different basis functions that lead to the characterization and localization features in both the space and transform domains at the same time. Furthermore, they offer the means of image representation that facilitates multiresolution image analysis. In fact, provided that appropriate functions are chosen, the wavelet transform domain representation, like the transform coefficient image, will offer spatial correspondence of image features in orientation, location, and scale. This makes for a more sensitive and selective image feature detection and enhancement process [4].

Figure 1. (a) A metaphase cell spread; (b) A karyotype of the chromosomes in (a).

The wavelet packet transform is an expanded basis function system that allows high resolution decomposition of a signal. These are computed by iterating not only down the low pass scaling function branch of Mallat’s Discrete Wavelet Transform (DWT ) logarithm tree, but also down the high pass wavelet branch [5]. Thus, the wavelet packet transform offers more functions for signal analysis than the wavelet transform alone. Compared with the wavelet transform, the wavelet packet transform uses only full-width functions that are all orthogonal; the first have zero area. These are intuitively more satisfying weighting functions than the ever-narrowing wavelet basis functions.

By taking advantage of the wavelet transforms described above, researchers have set out to explore the use of wavelet transforms for chromosome image analysis, specifically in enhancement and classification. Meanwhile, algorithms using wavelet transforms have been proposed for M-FISH images. This paper is focused on these proposed algorithms; their principle and realization are analyzed in detail.
II. BASIC THEORY OF WAVELET TRANSFORMS

The word “wavelet” has been used for decades in digital signal processing and in the exploration geophysics [6]. Wavelet theory is applicable to several subjects. All wavelet transforms may be considered forms of time frequency representation for continuous-time (analog) signals, and are thus related to harmonic analysis. Almost all useful discrete wavelet transforms (DWT) use discrete-time filter banks. These filter banks are called the wavelet and scaling coefficients in wavelets nomenclature. These filter banks may contain either finite impulse response (FIR) filters or infinite impulse response (IIR) filters. The wavelets forming a continuous wavelet transform (CWT) are subject to the Uncertainty Principle of Fourier analysis Respective Sampling Theory. Given a signal with some event in it, one cannot assign simultaneously an exact time and frequency response scale to that event. The product of the uncertainties of time and frequency response scale has a lower bound. Thus, in the scaleogram of a continuous wavelet transform of this signal, such an event marks an entire region in the time-scale plane, instead of just one point. Also, discrete wavelet bases may be considered in the context of other forms of the uncertainty principle. Wavelet transforms are broadly divided into three classes: continuous, discrete, and multiresolution-based [7].

III. ENHANCEMENT ALGORITHMS USING WAVELET TRANSFORMS

Chromosome images are acquired by microscopic imaging of metaphase or prophase cells on specimen slides. Digitized chromosome images usually are poor in quality, lacking contrast and shading. This is due to deficiencies in focus, lighting, and specimen staining. These images are hardly usable for visualization and karyotyping purposes. Computer processing and enhancement of chromosome images can largely assuage this problem [4]. The purpose of this preprocess is to improve image contrast, reduce noise, and correct for rotation, and overlapping of chromosomes.

One study described a new technique developed for human chromosome image enhancement based on the cubic-spline wavelet transform and multiresolution image analysis [4]. The cubic spline basis functions bore a remarkable shape resemblance to chromosome bands. This property was very desirable, since the transform represents the chromosome as a weighted sum of basis functions. This enabled the transform to decompose the chromosome into band-like components in the transform domain. And, by virtue of the space-frequency localization property and of the multiresolution representation capability of wavelet transforms, the image enhancement of the chromosome band patterns could be readily achieved by analyzing and selectively processing the wavelet transform coefficient image at the designated location, orientation, and scale. The experimental results showed that this technique improves image enhancement, allowing it to assist the evaluation of chromosome abnormalities in clinical samples.

Another study applied a class of differential wavelets to chromosome image enhancement [8]. The proposed differential filters facilitate the extraction of multiscale geometric features of chromosome images. Moreover, desirable fast computation can occur. This algorithm improves the salient features of chromosome images (including the band patterns), facilitates the measurement of correlation of image features in the transform domain, and provides high-frequency edge information along horizontal, vertical and diagonal directions. In order to objectively compare the performance of the proposed algorithm against several existing image-enhancement techniques, a quantitative criteria has been adopted to evaluate the results. The experimental results indicate that the proposed method consistently outperforms existing techniques.

One investigation described a novel method of enhancing image geometric features [9]. This method used the oriented wavelets introduced in the researcher’s early work, which were derived from isotropic Laplacian-like filters [10]. The oriented wavelets were used to circumvent the poor directional selectivity of the conventional 2-D wavelet transform. Because of the directional wavelet decomposition, the directional and scale correlation information could enhance the salient structures in images. Then, a pair of quadrature filters were used to remove the phase dependence. The preliminary results showed that this method significantly outperforms other conventional techniques in terms of improving the image contrast.

How to evaluate the effect of these different enhancement algorithms is a difficult yet important problem. One solution is to evaluate the performance difference of chromosome classification before and after enhancement, given that the same feature measurement and classifier methods are used in the comparison.

One research quantitatively examined the effect of image enhancement on the performance of chromosome classification using the above idea [11]. The experiments were conducted on a test set of chromosome images before and after enhancement. The test results show that the wavelet-based image enhancement technique [12] substantially reduces the error rate of chromosome classification. It can be learned from this research that proper image enhancement leads to significant improvements in recognition accuracy.

IV. CLASSIFICATION ALGORITHMS USING WAVELET TRANSFORMS

Chromosome karyotyping refers to the classification and subsequent formatted display of the chromosomes found in a cell spread. A karyotype is required to assign each chromosome to one of 24 classes (22 autosomes and two sex chromosomes). In order to improve the performance of automated chromosome classification, artificial intelligence and machine learning methods have been widely used. Among them, artificial neural networks (ANN) and wavelet transform algorithms are the most popular tools.

In one study, researchers employed wavelet packets as basis function sets to compute chromosome band pattern features [13]. A total of 28 wavelet packet basis function sets were evaluated, including the well-known Haar and
Daubechies’4 and Daubechies’6 wavelet packets. The experiments were conducted on two benchmark chromosome datasets (“Copenhagen” chromosomes and “Genzyme” set). The experimental results were then compared with the results of the current best-performing Weighted Density Distribution (WDD) method. From those results, it could be found that, on both tested datasets, the best-performing wavelet packet came very close to the WDD approach. The Haar, D4, and D6 however, were further behind on the accuracy scale. It also could be found that more ragged wavelet packet basis functions led to worse classification accuracy.

Another study proposed a classifier based on a wavelet neural network (WNN), combining the wavelet into a neural network for classification of chromosomes in group E (chromosomes 16, 17 and 18) [14]. According to the nonlinear characteristic of the chromosomes, WNN classified the chromosomes with high classification accuracy and high training speed. The features examined in the experiment were the length, the CI, the normalized average gray density (NAGD), the position, the width, and the intensity. These factors were used as the network inputs. Then, a three layer WNN was employed for classification of chromosomes in group E. The outputs were three classes. According to the experimental results, the WNN achieved high accuracy with minimum training time, which makes it suitable for real-time chromosome classification in the laboratory.

One research investigated and compared the efficacy of the wavelet and the Fourier descriptors in neural networks used for chromosome classification [15]. In that study, the back propagation (BP) neural network architecture was used. Firstly, the density profile of each chromosome was computed. Then, the discrete wavelet transform (4-coefficient Daubechies wavelet basis) and discrete Fourier transform were applied to the density profile. After that, the transformed densitometric signals were equally sampled and used as analytic features. The testing results demonstrated that using Fourier transform based features could achieve a 2.8% higher accuracy compared with using wavelet transform based features. But, the wavelet processing simplified the classifier, meaning that there was less time to process the data. And, the wavelet networks also converged more quickly than the Fourier networks, so wavelet networks are more efficient than Fourier networks.

A method for chromosome classification based on the chromosome shape was proposed [16]. First, the original image was filtered by the median filter and binarized by the Otsu method. Second, the zero-crossing technique was applied to detect the edge of the chromosome. After that, the signature of the contour was produced. Third, the wavelet packet transform was applied to decompose the signature. Finally, the chromosomes were classified by comparing the coefficients of the best tree of the wavelet packet transform correspondent to the signature of chromosomes. The results obtained showed that the proposed method provides the classification of the chromosomes.

In one research, an automatic procedure for chromosome image analysis was introduced [17]. First, the input chromosome images were pre-processed, reducing the noise of the image and enhancing the band features. Then, the pale-path algorithm was proposed to segment touching and overlapping chromosomes, obtaining the individual chromosomes. After that, the middle point algorithm was used to extract the medial axis. The algorithm based on the multiscale wavelets B-spline was applied to enhance the chromosome band. Then the enhanced band could be extracted by the average gray profile, the gradient profile, and the shape profile. Finally, the multilayer classifier was used to classify the chromosome pattern calculated by the weighted density distribution algorithm. The experiment results demonstrated that the algorithms perform well.

One shortcoming of the above algorithms is that they cannot work on bent chromosomes. So, an effective algorithm for chromosome image processing for straightening the curved chromosomes was presented [18]. After pre-processing and histogram modification of the input image, the binary image was obtained by thresholding the processed image. Then, the image was rotated and the projection vectors for each image were computed. After analyzing the projection vectors, the bending axis of the chromosome was determined. Finally the chromosome could be straightened. When this algorithm was applied to real chromosome images, all of the highly bent or curved chromosomes within the image dataset could be straightened. In order to evaluate the effectiveness of this algorithm, the straightened chromosomes were used in a classification procedure. The wavelet transform domain features were extracted and used in an MLP structure for this purpose and a classification rate of 95.3% was obtained.

An automatic classification of the human chromosome method was proposed [19]. This method was based on the linear discriminant analysis of the wavelet domain features for discriminating both normal and automatically straightened chromosomes in group E. It demonstrated the excellent performance of the straightening algorithm [20] in the automatic classification of human chromosomes. First, the DWT was applied to extract a set of features from the Density Profile (DP) of the chromosome. Then, a feature reduction mapped the wavelet coefficients into a new feature set by using Linear Discriminant Analysis (LDA). Finally, the above feature set was added by the Centeromeric Index (CI) and the Relative Length (RL) of the chromosome. After that, the final feature vector was produced. The performance of the proposed feature vector was compared to those of other two feature sets obtained by DCT and PCA. The experiment results showed that the WT-LDA based feature set outperformed the other two feature sets with a feature vector size of almost 50% of the size of the other two feature vectors.

In another study, a hybrid intelligent karyotyping system was introduced [21]. The most important characteristic of this system was the ability to detect chromosome aberrations. The most important difference between this system and other existing ones was that its feature extraction phase used wavelet transform techniques. Throughout this system, the wavelet transform techniques were applied on
the gray level profiles of chromosomes to represent the band patterns of chromosomes.

V. WAVELET TRANSFORMS USED FOR M-FISH IMAGES

Multi-color fluorescence in situ hybridization (M-FISH) imaging is a recently developed molecular cytogenetic technique for the rapid visualization of genomic aberrations at the subcellular level. This genetic imaging technique is based on the simultaneous hybridization of a 24-chromosome-specific probe pool [22, 23]. The technique is also called the color karyotyping.

In one study, a new wavelet transform for multi-color image representation was proposed [24]. In this method, the differential wavelets were taken as the derivatives of B-splines. The wavelet transforms were defined as the convolutions with the dilations of these wavelets at different scales. This method effectively reduced redundancy between multi-color images. The testing on a real world M-FISH image data has demonstrated that this method can improve the accuracy of pixel-by-pixel classification of chromosomes from M-FISH imaging data.

M-FISH technology has been used more often for the diagnosis of genomic abnormalities in the field of cancer cytogenetics. As the use of digital media becomes more commonplace for cytogenetic image storage, archiving, and communication, efficient compression techniques are in high demand to accommodate the explosive growth of the digital image data and to reduce storage and transfer costs [25].

One study presented the embedded M-FISH image coding (EMIC), where the foreground objects (chromosomes) and the background objects (images) were coded separately [25]. First, the critically sampled integer wavelet transforms were applied to both the foreground and the background of the image. Then, object-based bit-plane coding was used to compress each object and generate separate embedded bit streams, allowing for continuous lossy-to-lossless compression of the foreground and the background. A method of designing an optimal context model was proposed for the efficient arithmetic coding of bit planes. This method specifically exploited the statistical characteristics of M-FISH images in the wavelet domain. The experiment results showed that EMIC performs significantly better than several popular coding techniques, including LZW and JPEG-LS.

VI. CONCLUSIONS

Since karyotyping is a time-consuming procedure, computer-based classifiers have been proposed. A large number of studies have been conducted to develop computer-assisted chromosome detection and classification systems, as well as to evaluate and improve their performance. M-FISH imaging is a recently established molecular cytogenetic technique for the rapid visualization of genomic aberrations at the subcellular level. Wavelet transforms offer the means of image representation facilitating multiresolution image analysis. Researchers have set out to explore the use of wavelet transforms for chromosome image analysis. This paper has discussed the algorithms for chromosome enhancement and classification using wavelet transforms. Meanwhile, the algorithms based on the wavelet transforms for M-FISH images have also been analyzed. The principles and results of these algorithms have been described in detail.

ACKNOWLEDGMENT

This work was financially supported by the National Natural Science Foundation of China (No. 61163050, No. 61472137), NCST Program (No. 3142013098, No. 3142014125), the Higher school science research project of Hebei Province (Z2014038) and the Fundamental Research Funds for the Central Universities (3142011047 and 3142014007).

REFERENCES


