Chinese Chemical Letters Vol. 14, No. 11, pp 1193 – 1195, 2003
http://www.imm.ac.cn/journal/ccl.html

Application of Embedded Zerotree Wavelet to the Compression of Infrared Spectra

Meng Long LI[1]*, Hua Yi QI[1], Fu Sheng NIE[2], Zhi Ning WEN[1], Bin KANG[1]

[1] College of Chemistry, Sichuan University, Chengdu 610064
[2] CODI, 27 E. Central Avenue, Suite I1, Paoli, PA 19301,USA

Abstract: In this paper the embedded zerotree wavelet (EZW) method and Huffman coding are proposed to compress infrared (IR) spectra. We found that this technique is much better than others in terms of efficiently coding wavelet coefficients because the zerotree quantization is an effective way of exploiting the self-similarities of wavelet coefficients at various resolutions.

Keywords: Embedded zerotree wavelet, compress, infrared spectra, wavelet transform.

The wavelet transform (WT) has been developed rapidly in recent years. It plays a significant role in signal processing. From 1989 onwards, the WT has been applied in chemical studies owing to its efficiency, available large number of basis functions and high speed in data treatment[1, 2]. IR spectroscopy has been found widespread use in identifying and characterizing chemicals. IR spectral libraries, such as the Aldrich Library, have been established for this purpose. In general, the more information stored within a spectrum, the better for searching the database is. However, as increase the size of library, more space and longer time for searching the library were required.

To tackle the problem, researchers may use several approaches. One is to compress the spectrum to reduce the required storage space. The WT has been proposed for this purpose. But, how to get the most efficient compression, *e.g.* how to process these wavelet coefficients efficiently is remained to be solved. In the previous works most researchers often adopted discarding small-amplitude values directly without considering the self-similarities of wavelet coefficients at various resolutions[3- 7]. In this work, we applied a different WT procedure. The embedded zerotree wavelet algorithm is used in analytical chemistry for the first time. This method is initially proposed by Shapiro to process images[8].

The EZW encoder is based on progressive encoding to compress a spectrum into a bit stream with increasing accuracy. It means that when more bits are added to the stream, the well-decoded spectrum will contain more details. Every added digit increases the accuracy somehow and it can be stopped at any accuracy we like. Furthermore this algorithm requires no training, no pre-stored table or codebooks, no prior knowledge of the spectral source. The original spectrum was decomposed to six successive scales by

* E-mail: cic@scu.edu.cn

discrete wavelet transform (DWT). After decomposition, EZW encoding method was applied to scan all the wavelet coefficients. In this way, the coefficients were recorded by some special symbols and encoded with code words by Huffman coding. The procedure has been sketched in **Figure 1** and documented in detail in **Figure 2** and **Figure 3**.

Figure 1 The procedure of compression with embedded zerotree wavelet

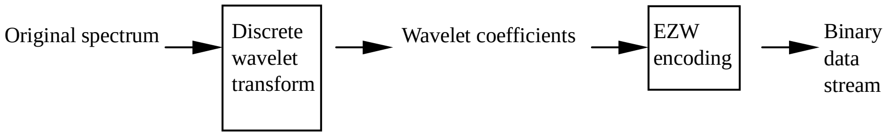

Figure 2 The relations between wavelet coefficients at different scales

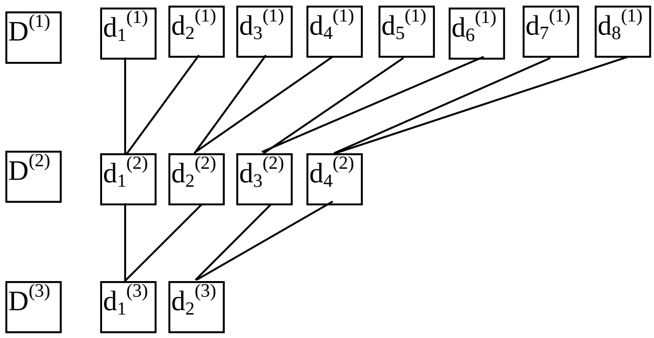

Figure 3 How to scan them (1) and the results of using zerotree symbols (T) in the coding process (2).

1
a scan order

2
standard:
HH LH LL LL LH HL LL
using zerotrees:
HH TH LT LH HL

The coefficient which is higher than the threshold (H),
The coefficient which is below the threshold.(L),
The zerotree symbol which replaces some L(T)

In this work, we use the biorthogonal wavelet series of bior 1.1, 2.8, 3.7, 4.4 and 6.8 respectively. Comparing to the direct threshold method in the previous works, it was found that our method gives an average compression ratio (CR) improvement by 36.1% and the biggest improvement by 68.1%. By the way, it is commonsense that the original data size will influence on compression ratio. Under the same conditions, the larger the original data number are, the higher the compression ratio is. In our experiment the data number was 460, it is often larger than in other works. So it is expected that a higher CR can be achieved if we compress the IR spectrum with data length longer than 460.

In conclusion, this kind of technique takes advantages of not only the character of wavelet coefficients that there are many small-amplitude values in them, but also the self-similarities of wavelet coefficients at various resolutions. The results showed that the performance of the proposed compression method is much better than those suggested in the previous studies[9, 10].

Acknowledgments

This work was partly supported by the National Natural Science Foundation of China (No. 29877016).

References

1. A.K.M. Leung, F.T. Chau, J.B. Gao, *Chemom. Intell. Lab. Syst.*, **1998**, *43*, 165.
2. X.G. Shao, C.Y. Pang, L. Sun, *Progress in Chemistry (in Chinese)*, **2000**, *12*, 233.
3. U. Depczynski, K. Jetter, K. Molt, A. Niem ller, *Chemom. Intell. Lab. Syst.*,**1999**, *49*, 151.
4. V.J. Barclay, R.F. Bonner, *Anal. Chem*., **1997**, *69*, 78.
5. A.K.M. Leung, F.T. Chau, J.B. Gao, T.M. Shih, *Chemom. Intell. Lab. Syst.*, **1998**, *43*, 69.
6. F.T. Chau, T.M. Shih, J.B. Gao, C.K. Chan, *Appl. Spectrosc*. **1996**, *50* , 339.
7. R.R. Coifman, Y. Meyer, S. Quake, M.V. Wickerhauser, in: J.S. Byrnes, K.A. Hargreaves, K.Berry (Eds), *Wavelets and their Applications*, Kluwer Academic Publishers, The Netherlands, **1994**, pp. 363.
8. J.M. Shapiro, *IEEE Trans. Signal Processing*, **1993**, *41* 3445.
9. M.L. Li, B. Kang, H.Y. Qi, Z.N. Wen *Chemistry Online*, **2002**, *65*, w71.
10. M.L. Li, B. Kang, H.Y. Qi, Z.N. Wen *Chem. J. Chin. Univ.*, **2002**, *23*, 1281.

Received 2 September, 2002