National Taiwan University
Graduate Institute of Communication Engineering

Time Frequency Analysis and Wavelet Transform
Term Paper

Introduction to Medical Image Compression Using Wavelet Transform

李自恒 **(Tzu-Heng Henry Lee) R96942133**

Instructor: Jian-Jiun Ding
Submitted: Dec. 31, 2007

Abstract

Compressions based on wavelet transform are the state-of-the-art compression technique used in medical image compression. For medical images it is critical to produce high compression performance while minimizing the amount of image data so the data can be stored economically. The wavelet-based compression scheme contains transformation, quantization, and lossless entropy coding. For the transformation stage, discrete wavelet transform and lifting schemes are introduced. The paper is concluded by discussing the applications of the wavelet-based image compression on medical images and radiologic practice.

1. Introduction

Modern radiology techniques provide crucial medical information for radiologists to diagnose diseases and determine appropriate treatments. Such information must be acquired through medical imaging (MI) processes. Since more and more medical images are in digital format, more economical and effective data compression technologies are required to minimize mass volume of digital image data produced in the hospitals [1], [2], [3].

Typically, compression scheme can be categorized into two major categories: lossless and lossy compressions. Lossless image compression can be achieved if the original input image can be perfectly recovered from the compressed data while lossy image compression cannot regenerate the original image data. Lossy image compression, however, is able to maintain most details of the original image that is useful for diagnosis. The precise detail preservation of an image is not usually strictly required because the degraded part of the image is often not visible to a human observer. But the lossy image compression is not very commonly used in clinical practice and diagnosis because even with a slight data loss, it is possible that physicians and radiologists missed the critical diagnostic information that could be a decisive element for the diagnosis of a patient and the following treatment [1], [2]. Lossless compression scheme cannot achieve a compression ratio that is greater than 4:1. In contrast, lossy compression can well preserve diagnostic quality at a compression ratio up to 100:1 [1]. It is obvious that the ability of lossy compression to obtain compact representation of the image is more advantageous with a tradeoff in some data loss.

Medical image compression based on wavelet decomposition has become a state-of-the-art compression technology since it can produce notably better medical image results compared to the compression results that are generated by Fourier transform based methods such as the discrete cosine transform (DCT) used by JPEG [2]. This paper provides an overview of the image decomposition technique using wavelet transforms, quantization algorithm based on embedded zerotree wavelet (EZW), and a lossless entropy encoder. The aim of this paper is to introduce how the wavelet compression techniques are applied in medical image compression in order to enhance the accuracy in medical diagnosis.

2. Compression Scheme Overview

In general, there are three essential stages in a transform-based image compression system: transformation, quantization, and lossless entropy coding. Fig. 1 depicts the encoding and decoding processes in which the reversed stages are performed to compose a decoder. The only different part in the decoding process is that the de-quantization takes place and it is followed by an inverse transform in order to approximate the original image. The purpose of transformation stage is to convert the image into a transformed domain in which correlation and entropy can be lower and the energy can be concentrated in a small part of the transformed image [1]. This energy packing property is illustrated in Fig. 2. Note that the vertical axis represents the number of pixels and the horizontal axis represents the pixel value. The narrow peak configuration as shown in Fig. 2(b) makes the transformed image easier to compress [3]. Quantization stage results in loss of data because it reduces the number of bits of the transform coefficients. Coefficients that do not make significant contributions to the total energy or visual appearance of the image are represented with a small number of bits or discarded while the coefficients in the opposite case are quantized in a finer fashion [1], [2]. Such operations reduce the visual redundancies of the input image [4]. The entropy coding takes place at the end of the whole encoding process. It assigns the fewest bit code words to the most frequently occurring output values and most bit code words to the unlikely outputs. This reduces the coding redundancy and thus reduces the size of the resulting bit-stream [1], [4].

3. Transformation

In contrast to image compression using discrete cosine transform (DCT) which is proved to be poor in frequency localization due to the inadequate basis window, discrete wavelet transform (DWT) has a better way to resolve the problem by trading off spatial or time resolution for frequency resolution. The DWT is based on a theory

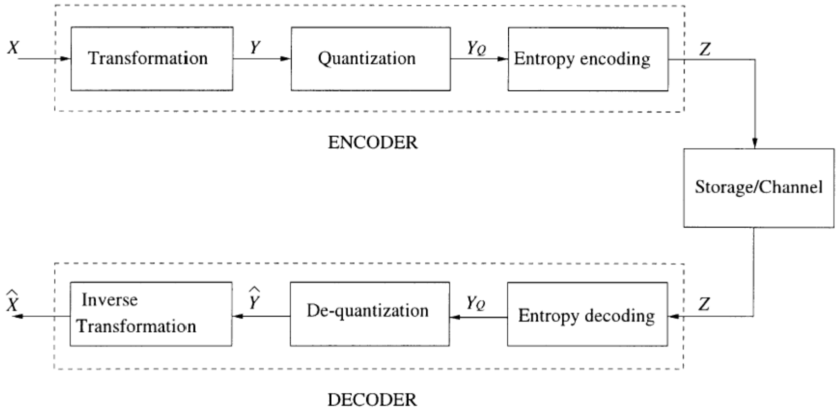
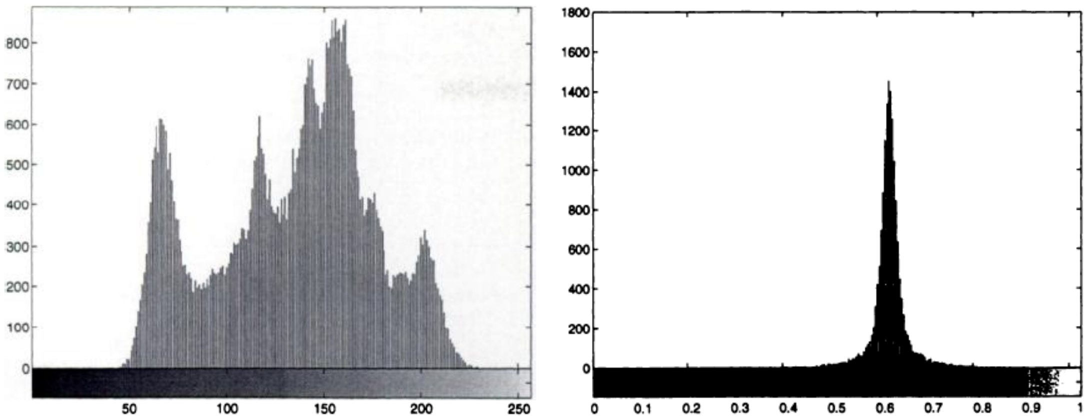


Fig. 1 Block diagram of the general compression and decompression processes [1].



(a) (b)

Fig. 2 Pixel values extracted from a chest CT scan in histograms before and after the wavelet transform. (b) clearly shows the energy distribution is concentrated after wavelet transform [2].

of multiresolution analysis (MRA) [1]. Detail on multiresolution analysis is not further elaborated in this paper.

3.1 The Discrete Wavelet Transform

The best way to describe discrete wavelet transform is through a series of cascaded filters. We first consider the FIR-based discrete transform. The input image x is fed into a low-pass filter $\tilde{h}$ and a high-pass filter $\tilde{g}$ separately. The outputs of the two filters are then subsampled. The resulting low-pass subband y_L and high-pass subband y_H are shown in Fig. 3. The original signal can be reconstructed by synthesis

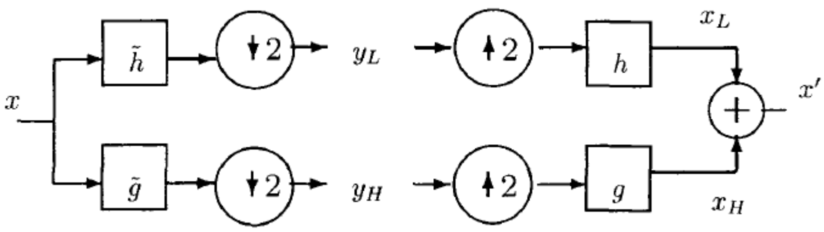


Fig. 3 DWT analysis and synthesis system [5].

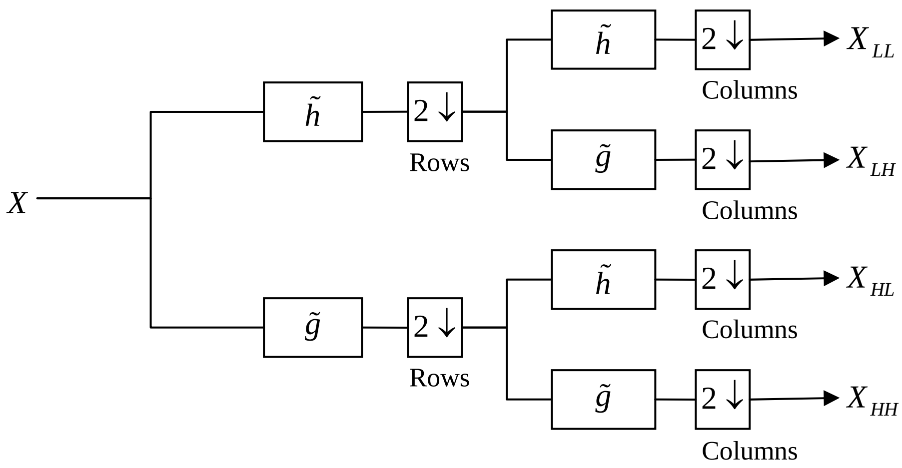


Fig. 4 The 2-D DWT analysis filter bank [1].

filters h and g which take the upsampled y_L and y_H as inputs [5]. An analysis and synthesis system has the perfect reconstruction property if and only if $x' = x$ [1]. The mathematical representations [5] of y_L and y_H can be defined as

$$y_L(n) = \sum_{i=0}^{\tau_L - 1} \tilde{h}(i) x(2n - i) \quad (1)$$

$$y_H(n) = \sum_{i=0}^{\tau_H - 1} \tilde{g}(i) x(2n - i) \quad (2)$$

where τ_L and τ_H are the lengths of $\tilde{h}$ and $\tilde{g}$ respectively. For a two dimensional image, the DWT have to be extended to the 2D case. We suppose the image to be compressed has a dimension of M rows by N columns. The approach of the 2D implementation of the DWT is to perform the one dimensional DWT in row direction and it is followed by a one dimensional DWT in column direction. This decomposition technique is shown in Fig. 4. A two dimensional row and column computation of DWT is depicted in Fig. 5. In the figure, LL is a coarser version of the original image and it contains the approximation information which is in low frequency. LH, HL, and HH are the high-frequency subband containing the detail

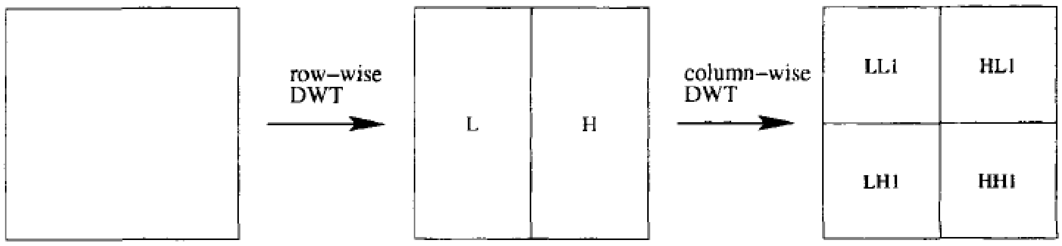


Fig. 5 The two dimensional row and column computation of DWT [5].

Fig. 6 The (Left) second and (Right) third level row and column decomposition [5].

information [5]. Further computations of DWT can be performed as the level of decomposition increases. This concept is also illustrated in Fig. 6. In Fig. 6, the second and third level decompositions based on the principle of multiresolution analysis show that the LL1 subband shown in Fig. 5 is decomposed into four smaller subbands: LL2, HL2, LH2, and HH2.

3.2 Lifting Scheme

Direct discrete wavelet transform implementation is theoretical invertible. However, due to the finite register length of the computer system, inversion errors could occur and it would result in unsuccessful image reconstruction. In practical cases, the wavelet coefficients will be rounded to the nearest integer in the discrete transformation stage. This makes the lossless compression impossible [1]. An improved implementation called lifting-based wavelet transform which is based on the wavelet theory is proposed and it requires significantly fewer arithmetic computations and memory compared to the convolution based discrete wavelet transform. The lifting-based DWT scheme breaks up the high-pass and low-pass wavelet filters into a sequence of smaller filters. These decomposed filters are then converted into a sequence of upper and lower triangular filters [5]. The derivations of the triangular matrices are not presented in this paper.

In order to achieve perfect reconstruction of a signal, the filters shown in Fig. 3 must satisfy the following conditions:

$$\begin{cases} h(z)\tilde{h}(z^{-1}) + g(z)\tilde{g}(z^{-1}) = 2 \\ h(z)\tilde{h}(-z^{-1}) + g(z)\tilde{g}(-z^{-1}) = 0 \end{cases} \tag{3}$$

In the lifting scheme, the impulse response coefficients *h* and *g* are expressed in Laurent polynomial with the aid of Z-transform. For instance, the Laurent polynomial representation of filter *h* can be defined as

$$h(z) = \sum_{i=m}^{n} h_i z^{-i} \tag{4}$$

where *m* and *n* are positive integers. The analysis and synthesis filters as shown in Fig. 3 are further decomposed into the polyphase representations which are expressed as

$$h(z) = h_e(z^2) + z^{-1}h_o(z^2) \tag{5}$$

$$g(z) = g_e(z^2) + z^{-1}g_o(z^2) \tag{6}$$

$$\tilde{h}(z) = \tilde{h}_e(z^2) + z^{-1}\tilde{h}_o(z^2) \tag{7}$$

$$\tilde{g}(z) = \tilde{g}_e(z^2) + z^{-1}\tilde{g}_o(z^2) \tag{8}$$

The two polyphase matrices of the filter *h* is defined as

$$P(z) = \begin{bmatrix} h_e(z) & g_e(z) \\ h_o(z) & g_o(z) \end{bmatrix}, \qquad \tilde{P}(z) = \begin{bmatrix} \tilde{h}_e(z) & \tilde{g}_e(z) \\ \tilde{h}_o(z) & \tilde{g}_o(z) \end{bmatrix} \tag{9}$$

The parameter *z* is used since the polyphase representations are derived using Z-transform and the subscript *e* and *o* denote the even and odd sub-components of the filters which are split into subsequences. The purpose of the polyphase representation is to reduce the computation time. The perfect reconstruction is ensured only when the following relation is true [1], [5]:

$$P(z)\tilde{P}(z^{-1})^T = I \tag{10}$$

where *I* is a 2 by 2 identity matrix. Now the wavelet transform can be expressed using the polyphase matrix for forward discrete wavelet transform [5]

$$\begin{bmatrix} y_L(z) \\ y_H(z) \end{bmatrix} = \tilde{P}(z) \begin{bmatrix} x_e(z) \\ z^{-1}x_o(z) \end{bmatrix} \tag{11}$$

and the inverse discrete wavelet transform [5]

$$\begin{bmatrix} x_e(z) \\ z^{-1}x_o(z) \end{bmatrix} = P(z) \begin{bmatrix} y_L(z) \\ y_H(z) \end{bmatrix} \tag{12}$$

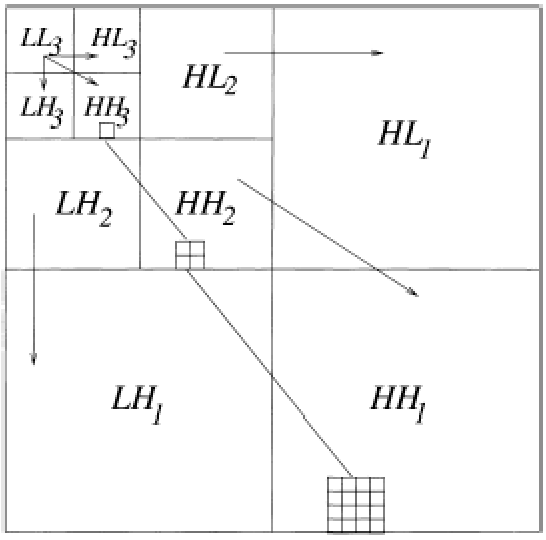


Fig. 7 A three level wavelet decomposition of image spectrum [1].

Finally, the upper and lower triangular matrices can be obtained by the lifting factorization process. The lifting sequences are generated by employing Euclidean algorithm which factorizes the polyphase matrix for a filter pair [5].

4. Quantization

The goal of quantization stage is to reduce the values of the transformed coefficients in order to reduce the precision of the subbands and achieve the compression [2]. As it is mentioned in section 2, the quantization stage is the main cause of the information loss in the encoder. Therefore, a quantization process is only employed in lossy compression and it is not performed in lossless compression [5]. The embedded zerotree wavelet (EZW) is an effective algorithm employed in this stage. At a given compression ratio in bit rate, EZW is able to achieve the best image quality and encode the image so that all lower bit rate encodings are embedded at the beginning of the final bit-stream. In the EZW's algorithm, the information on which the coefficients are significant is generated and then encoded via quantization. The significance map determines whether a DWT coefficient is to be quantized as zero or not. A DWT coefficient is considered insignificant with respect to a given threshold T if $|x| \leq T$. Otherwise a coefficient is called significant. Since the wavelet decomposition has the hierarchical structure in which each coefficient can be related to a set of coefficients that is at the next finer resolution level, a tree structure depicted in Fig. 7 can be defined as the concept of descendants and ancestors. Given a threshold T to determine whether or not a coefficient is significant, a coefficient x is said to be an element of a zerotree root (ZRT) for the threshold *T* if itself and all of its descendents are insignificant with respect to the threshold *T*. For the case which not all the descendants are insignificant, the coefficients are encoded as isolated zero (IZ). For encoding a significant coefficient, the symbol POS and NEG are used. Therefore, given a threshold *T*, the wavelet coefficients can be represented by the four symbols: zerotree root (ZRT), isolated zero (IZ), positive significant (POS) and negative significant (NEG) [1].

5. Entropy Coding

The output symbol stream is an input to an entropy encoder to complete the last stage of the compression without adding distortion. The lossless entropy encoding process replaces the symbol stream produced in the quantization stage with a sequence of binary codewords which is called a bit stream [1], [2]. The probability of the corresponding symbol is proportional to the length of a codeword. The smallest possible number of bits that is required to represent a symbol sequence can be defined as the entropy of the symbol source:

$$H = -\sum_i p_i \log_2 p_i \qquad (13)$$

Here the p_i is the probability of the *i*th symbol. In the optimal case, the sum of the probability $\sum_i p_i$ would be equaled to 1 and the *i*th symbol would be $-\log_2 p_i$ [1]. We can define the entropy as the expected length of binary code over all possible symbols.

In entropy encoding stage, source coding algorithms such as Huffman coding and arithmetic coding are often employed. Huffman coding is an optimal technique that is used to obtain the minimum average code length based on the associated probabilities which symbols might occur [1], [5]. In Huffman coding, shorter codewords are assigned to symbols that occur more frequently and longer codewords are assigned to symbols that occur less frequently [5]. The major drawback of employing Huffman coding is that the coding assigns an integer number of bits to every single symbol; this requires the probabilities to be powers of 2 in order to have the entropy bound. Arithmetic coding is another efficient method which can theoretically obtain the entropy lower bound even if the entropy bound is fractional. This is an approach which does not require a codeword to correspond to a symbol. In other words, a sequence of input symbols can be allocated with any real numbers

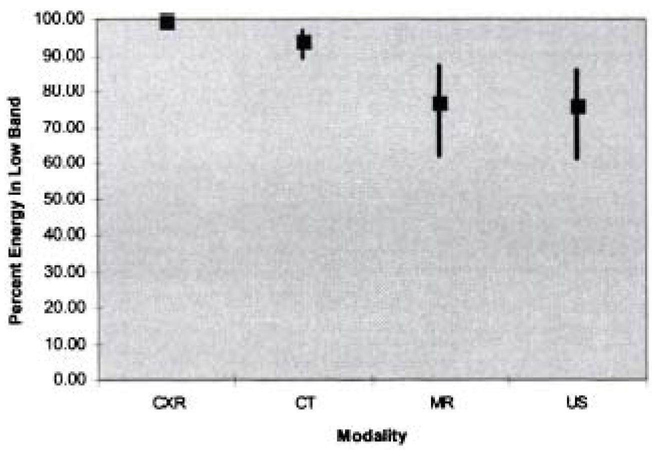


Fig. 8 Percentage energy in the lowest energy subband for 10 typical images from each of the chest radiography (CXR), computed tomography (CT), magnetic resonance (MR), and ultrasound (US) modalities [3].

between zero to one. Therefore, symbol can be represented by less than 1 bit [1], [5]. This is different from the Huffman coding and the other traditional encoding techniques in which each input symbol is replaced by a specific code with a certain integer number of bits. Longer symbol containing messages will result in the smaller interval between 0 and 1. The main idea of the arithmetic coding is to allow more probable symbols to reduce the interval less than the less probable symbols [5]. So the overall bits in the encoded message are fewer.

6. Discussions

The percentage of energy in the lowest frequency subband, as suggested in [3], can be used to indicate the compression tolerance for image. This idea is illustrated in Fig. 8. The figure demonstrates the percentage energy in the lowest energy subband for 10 typical images from each of the chest radiography (CXR), computed tomography (CT), magnetic resonance (MR), and ultrasound (US) modalities. From the figure, it is shown that the CXR is very tolerant of compression while the rest three modalities exhibit low tolerance to compression because considerable amount of the energies of CT, MR and US are concentrated in high frequency subbands. At low level of compression, it is likely that the image degradation could be imperceptible to human eyes. But at high level of compression, the error due to the loss of information during the compression process would become obvious [3]. Compression tolerance is crucial information for the digital compression of medical images since the overall wavelet compression process could cause loss of information in the compressed data. The compressed image quality (or the diagnostic quality) is exactly what the radiologists are concerned.

7. Conclusions

A basic introduction on the wavelet-based medical image compression is presented. Previous works demonstrated that compression of medical image data using irreversible wavelet transform appears to be a more effective approach to store and transmit radiologic images compared to other traditional Fourier-based compressions. Although using lossy (irreversible) compression leads to loss of some information that could be considered diagnostically insignificant, it can be commonly used to greatly reduce the cost of radiologic operation. For lossless compression, the lifting operation is another promising technique that is developed based on the theory of multiresolution analysis. The lifting realization lowers the complexity and theoretically provides perfect inversion characteristics. In general, the JPEG technique using DCT does not achieve compression ratios as high as those of most wavelet methods and wavelet compression certainly performs better than JPEG compression. With the techniques introduced in this paper, the diagnosis that replies on radiology seemed to be a lot more promising.

8. References

[1] E. Kofidisi, N.Kolokotronis, A. Vassilarakou, S. Theodoridis and D. Cavouras, "Wavelet-based medical image compression," *Future Generation Computer Systems*, vol. 15, no. 2, pp. 223-243, Mar. 1999.

[2] D. F. Schomer, A. A. Elekes, J. D. Hazle, J. C. Huffman, S. K. Thompson, C. K. Chui and W. A. Jr. Murphy, "Introduction to wavelet-based compression of medical images," *Radiographics*, vol. 18, no. 2, pp. 469-481, Mar. 1998.

[3] B. J. Erickson, A. Manduca, P. Palisson, K. R. Persons, F. IV. Earnest, V. Savcenko and N. J. Hangiandreou, "Wavelet compression of medical images," *Radiology*, vol. 206, no. 3, pp. 599-607, Mar. 1998.

[4] R. C. Gonzalez and R. C. Woods, "*Digital image processing*", Mass. Addison-Wesley, 1992.

[5] T. Acharya and A. K. Ray, *Image Processing: Principles and Applications*. Hoboken, NJ: John Wiley & Sons, 2005