Image Coding Using the Self-Similarity of Wavelet High-Frequency Components

S.Selvarajan
Dept. of Physical Science
Vavuniya Campus of the University of Jaffna
Station Road
Vavuniya 43000
Tel: 024-2220179
E-Mail:ksselvarajan@vavu.jfn.ac.lk

N.D.Kodikara
Dept. of Comm. & Media Tech.
University of Colombo
School of Computing
Colombo-03
Tel: 2591063
E-Mail:ndk@ucsc.cmb.ac.lk

Abstract

In this paper a novel approach Self-Similar Wavelet High-Frequency Components coding (SWHFC) to compress an image is proposed, based on the self-similarity of wavelet high-frequency components. The self-similarity of wavelet transformed image is explored by using the multiresolution concept of weighted finite automata (WFA). This scheme combines the multiresolution concept of WFA and wavelet multiresolution concept by introducing a predicative coefficients along with the each weights of WFA. Objective of this is to predict wavelet coefficients across scales, as to reduce the number of wavelet coefficients to be coded and the corresponding information of those.

Keywords: Image compression, wavelets, MRA, WFA, image smoothness, linear and nonlinear approximation

1. Introduction

In lossy image compression, two types of schemes are in use: fractal and transform coding. Fractal coding based on IFS, first proposed by Jacquin [7]. WFA is the generalization of IFS and proposed by Culik [1]. Transform coding scheme uses wavelet transform as an alternative to the DCT used earlier in JPEG.

In image processing most of the images are typically represents spatial trends, or areas of high statistical spatial correlation. However, anomalies, such as edges or object boundaries, take on perceptual significance that is far greater than their numerical energy contribution to an image. The basic idea of wavelet transform is to represent any arbitrary function $f$ as a superposition of wavelets. Any such superposition decomposes $f$ into different scale levels, where each level is then further decomposed with a resolution adapted to the level. One way to achieve such a decomposition writes $f$ as an integral over $a$ and $b$ of $\psi_{a,b}$ with appropriate weighting coefficients. Store only the relevant frequency components known as wavelet coefficients. A major difficulty is that fine detail coefficients representing possible anomalies constitute the large number of coefficients and therefore to make effective use of the multiresolution representation, much of the information is contained in representing the position of those few coefficients corresponding to significant anomalies.

The technique of this research allow coders to effectively use the power of multiresolution representation by efficiently representing the positions of the wavelet coefficients representing significant anomalies. A notable breakthrough was the introduction of Embedded Zero-tree Wavelet (EZW) coding by Shapiro [6]. In view of this context, it has been proposed to code natural images by nonlinear estimation of wavelet coefficients and combining both WFA and EZW by bounding the error incurred by quantizing wavelet transform coefficients and coding in WFA.

In this paper we present a recent result in combining those two, using the multiresolution properties of these two schemes. Culik [2] proposed a method to combine theses two scheme for fractal and smooth images, it reconstruct the image from the directly from WFA, where coefficients are coded. Our scheme differs from that, as it predict the coefficients by using the multiresolution concept of both and hierarchical nature of wavelet the transform. It is also applicable to natural images as well.

2. Weighted Finite Automata (WFA)

WFA is introduced by Culik as a device to compress images from the pixels grey-scale value. Later it was studied by Hafner [3]. The earlier is known as linear automata and
the later is known to be hierarchical because of its nature. We use the second one in our algorithm, this is somewhat similar to Laplacian Pyramidal coding [8].

Formal definition of the WFA is given [1], [3]. Wavelet transformed image represented in Mallat [4] form could be understandable as an independent image at different resolution in their corresponding orientation, i.e., horizontal, vertical and diagonal. These images (high-frequency components) are similar to each other by a the scaling factor with respect to their orientation. So that, coding the each highest-frequency component and low-frequency components are suffice to reconstruct the image by using the prediction across the scales of wavelet transform.

3. Image Compression

By an image, it means a digitized grey-scale picture, that $2^m \times 2^n$ pixels, $m \in \mathbb{Z}$ each of which takes a value $p_j$ such that, $0 \leq p_j \leq 2^m - 1$, where $j = (j_1, j_2)$, $j_1$ and $j_2$ are index of rows and columns respectively. In this analysis an image is described as a function $f$ on a unit square $I = [0, 1]^2$.

$$f(x) = p_j \quad \text{for} \quad \frac{j_1}{2^m} \leq x_1 \leq \frac{j_1 + 1}{2^m} \quad \text{and} \quad \frac{j_2}{2^m} \leq x_2 \leq \frac{j_2 + 1}{2^m}$$

where $x = (x_1, x_2)$ in $I$

By applying discrete transformation, e.g., DCT, Haar transform, we have

$$f = \sum_{j \geq 0, j=(j_1, j_2)} c_{j,k} \varphi_{j,k}$$

where $c_{j,k}$ are coefficients as the result of applying discrete transformation on the basis function $\varphi_{j,k}$.

Now the image compression problem viewed as the result of approximating $f$ by a second (compressed) function $\tilde{f}$, i.e., given the transform, the algorithm then calculates the quantized coefficients $\tilde{c}_{j,k}$ and the compressed function takes the form as:

$$\tilde{f} = \sum_{j \geq 0, j=(j_1, j_2)} \tilde{c}_{j,k} \varphi_{j,k}$$

4. Wavelet Representation of an Image

This section describes theoretical property of how an image is viewed as an image itself in wavelet transformation. Also, the principle underling behind the coding of wavelet coefficients in par with pixels being coded with WFA.

A digitized image is that the pixel value (observed) are samples, which depend on measuring device of the intensity field $F(x)$ for $x$ on the square $I = [0, 1]^2$. Pixel samples are well modeled by averaging the intensity function $F$ over all squares.

Assume that $2^{2m}$ pixel values $p_j$ are indexed by $j = (j_1, j_2)$, $0 \leq j_1, j_2 \leq 2^m$ of $2^{2m}$ rows and columns and that each measurement is the average value of $F$ on the subquadrant covered by that pixel. To fix this notation, the $j^{th}$ pixel covers the square $I_{j,m}$ with side length $2^{-m}$ and lower left corner at the point $j/2^m$. Denote the characteristic function of $I$ by $\chi = \chi_1$ and $L(I)$-normalized characteristic function of $I_{j,m}$ by $\chi_{j,m} = 2^m \chi_1(2^m \cdot j)$. Then the pixel value would be as:

$$p_j = 2^m \int \chi(2^m \cdot x - j) F(x) dx$$

$$= 2^{2m} < \chi_{j,m}, F >$$

The standard practice in wavelet-based image processing is to use the observed pixel value $p_j$ to create the function.

$$f_m = \sum_j p_j \chi(2^m \cdot x - j)$$

$$= \sum_j < \chi_{j,m}, F > \chi_{j,m}$$

which is known as observed image. If the wavelet expansion of the intensity field $F$ is

$$F = \sum_{0 \leq j, \psi} \sum_{0 \leq k, \psi} c_{j,k,\psi} \psi_{j,k}$$

then we have

$$f_m = \sum_{0 \leq j, \psi} \sum_{0 \leq k, \psi} c_{j,k,\psi} \psi_{j,k}$$

4.1. Multiresolution Concept

In computer vision, it is difficult to analyze the information content of an image directly from the grey-level intensity of the image pixels. The size of the neighborhood where the contrast is computed must be adapted to the size of the objects. The size define the resolution of reference for measuring the local variation of the image. Given a sequence of increasing resolutions $(r_j)$, the details of an image at the resolution $r_j$ are defined as the difference of information between its approximation at the resolution $r_j$.
and its approximation at the lower resolution \( r_{j-1} \).

Representing a function by a array of pixels smooths out details smaller than a pixel. But, because the scale of observation is arbitrary, one pixel can represent an area of any size. To avoid this problem, scale independent representation should be developed. This could be achieved through the use of multi resolution analysis. Intuitively, a MRA is a collection of subsets \( V_i \) of \( L_2, \ i \in \mathbb{Z} \). Each \( V_i \) contains all functions whose details smaller than \( (r_i)_{i \in \mathbb{Z}} \) is removed. Removal of details depends upon the particular type of MRA.

The Laplacian pyramid data structures suffer from the difficulty that data at separate levels are correlated. There is no clear model which handles this correlation. It is thus difficult to know whether a similarity between the image details at different resolutions is due to the property of the image itself or to the intrinsic redundancy of the representation. Further, this does not introduce any spatial orientation selectivity into the decomposition process.

Pyramidal implementation have been developed for computing the multiresolution transform based on convolution with quadrature mirror filters. The signal can be reconstructed by reversing the above process. A multiresolution transform also decomposes the signal into a set of frequency channels of constant bandwidth on logarithmic scale. It can be interpreted as a discrete wavelet transform.

4.2. Multiresolution Representation of an Image

For computational reason and self-similarity of the space-frequency plane of sub-bands, the Haar transform of the representation of \( f \) is considered. Let

\[
\Psi(x) = \begin{cases} 
-1 & 0 \leq x < 1/2 \\
1 & 1/2 \leq x < 1 
\end{cases}
\]

and

\[
\varphi(x) = 1 \quad 0 \leq x < 1
\]

The four basis function for the local Haar representation of functions for 2-D are:

\[
\psi^1(x, y) = \Psi(x)\psi(y) \quad \psi^2(x, y) = \Psi(x)\varphi(y) \\
\psi^3(x, y) = \varphi(x)\psi(y) \quad \psi^4(x, y) = \varphi(x)\varphi(y)
\]

By dilation and translation, we have \( \psi^{j0}_{j,k} = \psi^{j0}(2^j \cdot - j) \)

where each pixel \( p_{j,k} \), \( j \in \mathbb{Z}^2 \), is the average of the intensity \( F(x) \) on \( I_{j,m} \) which is defined to be the square of side length \( 2^{-m} \) and lower left corner located at \( j/2^m \). Consider the smoothness of any order \( 0 < \alpha < \beta \). The \( L_2(I), \ W_0^m(L_2(I)), 0 < \alpha < 1/2 \) and \( B_0^q(L_q(I)), 0 < \alpha < 1 \) and \( q = 2/(1 + \alpha) \), norms of \( f \) are bounded by the corresponding norms of \( F \). Therefore, linear approximation \( f \) of \( f \) is given by:

\[
f = \frac{c_{j,k-1}^{(1)}}{4}\psi_{j,k-1}^{(1)} + \ldots + \frac{c_{j,k-1}^{(4)}}{4}\psi_{j,k-1}^{(4)}
\]

where \( c_{j,k}^{(a)} \) is the grey-scale of a pixel

\[
\begin{pmatrix}
\frac{c_{j,k-1}^{(1)}}{4} & \frac{c_{j,k-1}^{(2)}}{4}
\end{pmatrix} = \begin{pmatrix}
-1 & -1 & 1 & 1 \\
1 & -1 & 1 & 1 \\
1 & 1 & 1 & 1 \\
1 & -1 & -1 & 1
\end{pmatrix} \begin{pmatrix}
P_{(2j_1,2j_2),k} \\
P_{(2j_1+1,2j_2+1),k} \\
P_{(2j_1+1,2j_2+1),k} \\
P_{(2j_1+1,2j_2+1),k}
\end{pmatrix}
\]

all the \( p_{j,k} \) are integers, then so are the \( c_{j,k-1}^{(a)} \). With this transform we have:

\[
f = \frac{1}{4} \sum_{k \geq 0} \sum_{j \in \mathbb{Z}} \sum_{k \geq 0} c_{j,k-1}^{(a)}\psi_{j,k-1}^{(a)}
\]

4.3. Similarities Across Scales

Wavelet coefficients associated with high-frequency components depends on a small rough region located around the edge and not on of overall smoothness of the picture. Conversely, smooth regions are less likely to be rippled due to nearby edge. The projection of a function \( f \) on two sub-spaces \( V_i \) and \( V_{i+1} \), similarities between them are observed. These two projections isolates two different frequency sub-bands. From the previous fact, if function \( A_{i} f \) is rough in some region, \( A_{i+1} f \) is also rough in the corresponding region. Our aim is to explore the nature of this similarity between the frequency sub-bands.

4.4. Linear and Nonlinear Approximation

Let \( f(x) = p_{j} \) for \( x \in I_{j,m} \) (support of the image \( I \) defined as earlier)

\[
F = \sum_{k \geq 0} \sum_{j \in \mathbb{Z}} c_{j,k} p_{j,k}
\]

\[
f_m = \sum_{0 \leq k \leq m} c_{j,k} p_{j,k}
\]

where each pixel \( p_{j,k} \), \( j \in \mathbb{Z}^2 \), is the average of the intensity \( F(x) \) on \( I_{j,m} \) which is defined to be the square of side length \( 2^{-m} \) and lower left corner located at \( j/2^m \). Consider the smoothness of any order \( 0 < \alpha < \beta \). The \( L_2(I), \ W_0^m(L_2(I)), 0 < \alpha < 1/2 \) and \( B_0^q(L_q(I)), 0 < \alpha < 1 \) and \( q = 2/(1 + \alpha) \), norms of \( f \) are bounded by the corresponding norms of \( F \). Therefore, linear approximation \( f \) of \( f \) is given by:
\[ f_k = \sum_{0 < \lambda < K} \sum_{j,k} c_{j,k,\lambda} \Psi_{j,k} \]  

where \( K \) is the smallest integer such that \( \lambda 2^{\beta K} \geq 1 \). This means, approximating all the coefficients \( c_{j,k,\lambda} \) with frequency less than \( 2^k, K \leq m \).

In nonlinear approximation, fix a number \( n \geq 0 \) and approximate \( f \) by
\[ \hat{f}_n = \sum_{\lambda \in \Lambda_0} c_{n,\lambda} \Psi_{\lambda} \]  

where \( \Lambda_0 \) is an arbitrary subset of \( \Lambda \) with \( 2^n \) elements. The best nonlinear approximation is obtained by taking \( \Lambda_0 \) as the set of the \( 2^n \) indices \( \lambda \) for which \(|c_\lambda|\) is the largest. With this process the image is smoothen, just as convolving with a smoothing kernel.

5. Wavelets, Quadtrees and WFA

Because of the dyadic nature of the wavelet decomposition, the 2-D wavelet transform is typically arranged in three subbands, corresponding to their orientation of the wavelet basis functions. Each subband can be organized into a quadtree, as described below.

In the quadtree interpretation of the 2-D wavelet transform, each node \( i \) is labeled with a wavelet coefficient \( c_{i,\lambda} \), where the corresponding wavelet basis function \( \psi_{\lambda} \) has approximate support on a square, dyadic block \( B_i \), in the image. The width of this block is given by \( M = 2^{-l}N \), where \( l \) is the depth of node \( i \) in the quadtree and \( N \) is the width in pixels of the square image (assumed to be the power of two).

Except at the finest level, each node has four children representing \( M/2 \times M/2 \) dyadic blocks that combine to tile the same \( M \times M \) image block as their parent. Objective is to code the finest level coefficients from which estimate the next lower high-frequency subband components, inorder to reduce the coding, position information of each wavelet coefficients.

Images are transformed using Haar basis orthogonal function. Each sub-band is assumed as the independent images and performed into horizontal, vertical and diagonal groups. The correlation between these groups of image components are very heigh, i.e., wavelet transform by Haar basis orthogonal function carries out the mean value among adjacent pixels can decompose into the space-frequency. Exactly, this process repeats the finite difference operation after averaging the adjacent pixels.

This could be seen as, higher frequency component of each frequency sub-band, wavelet coefficients in quad-tree form \((2 \times 2)\) having as the initial distribution and each edge having the weight \( w_{\lambda,n} \). It could be repeated for finite number of times to infer the wavelet coefficients in the corresponding lower frequency sub-bands. See figure 1.

5.1. Defining an Automata for Higher-frequency Components

It is noted high-frequency components of an image are similar across scales. It is obvious to define a WFA to represent such similarities. Since the WFA is the generalization of IFS. Also, wavelet transform consists hierarchical nature and multiresolution property. These could be incorporated in WFA. These properties ensure the multiresolution property of an image high frequency components, i.e., a function at resolution \( 2^k \) can be computed from the resolution \( 2^{k+1} \) by averaging the adjacent \( 2 \times 2 \) coefficients. On the basis of above concept define this representation as follows:

**Definition 1** Let \( c_{i,\psi}(2i+m,2j+n) \neq 0 \) be the coefficient of highest-frequency components and \( w(m,n) \) be the weight of each edge that correspond to next lower high-frequency component coefficient in the same orientation, for levels \( 0 < l \leq N \) and \( 0 \leq i,j \leq 2^l \), \( k \in Z \), then the coefficient of the next high-frequency component in the same orientation is estimated as:
\[ c_{i,\psi}(j^{l+1},j^{l+1}) = \sum_{m=0}^{3} \sum_{n=0}^{3} w(m,n) c_{i,\psi}(2i+m,2j+n) \]

where \( w(m,n) = 1/2 \) or \( 1 \) (scale invariant coefficient).

6. Zerotree for Smooth Regions

The smoothness of wavelet is often characterized by the number of vanishing moments. A function \( f \) defined over an interval \([a,b]\) is said to have \( n \) vanishing moments if and only if

![Figure 1: Coefficients construction.](image-url)
\[ \int_a^b f(x)x^i\,dx = 0 \quad \text{for} \quad i = 0, 1, \ldots, n - 1 \quad (14) \]

It refers to the decay of wavelet coefficients through the scales. Intuitively, we expect smooth image regions will have small wavelet coefficients.

Suppose node \( i \) has support on a dyadic image block \( B_i \), that is characterized as smooth. Because all nodes descending from \( i \) can also be characterized as smooth, this model assumes \( c_{i,j} = 0 \), as such corresponding coefficients in the next frequency subband component is zero. It is simply a fixed approximation to the wavelet coefficients. This tree-structured approximation is known as a zero-tree [6]. A combination of zerotree, nonlinear approximation and prediction of wavelet coefficients through the scale of high-frequency sub-band components are suffice for image coders.

A notable breakthrough in wavelet based coding was the introduction of Embedded Zerotree Wavelet (EZW) coding by Shapiro [6]. A significance map was defined as an indication of whether a particular coefficient was zero or nonzero (i.e., significant) relative to a quantization level. Defining a wavelet coefficient as insignificant with respect to a threshold \( T \) if \( |c_{i,j}| < T \), the EZW algorithm hypothesized that if a wavelet coefficient at coarse scale is insignificant with respect to a given threshold \( T \), then all wavelet coefficients of the same orientation in the same spatial location at finer scales are likely to be insignificant with respect to \( T \).

- if the corresponding coefficient in the next lower frequency sub-band is insignificant, then all the four coefficients are insignificant regardless of their values
- similarly, if the coefficients produce insignificant coefficient value to next lower frequency sub-band with our algorithm, then all four coefficients are insignificant

7. Coding and Decoding Algorithm

Images are transformed in three down sampling, using Haar basis orthogonal function. Haar basis function carries out the mean value among the adjacent pixels can decompose into the space-frequency. Exactly, this process repeats the finite difference operation after averaging the adjacent pixels. So that, it could be viewed as, higher-frequency components are in quad-tree form \( 2 \times 2 \) as the initial distribution and each having the weight \( w_{n,m} \) may be 1/2 or 1. Usually the weight is 1/2, in case of scale invariant coefficients the weight is 1. With this terminology, the corresponding next lower high-frequency sub-band coefficients could be inferred from (13). This process could be repeated for finite number of times to infer the coefficients of high-frequency sub-band components.

**Encoding Algorithm:**

1. Given image is transformed using Haar orthogonal basis filter in three dyadic levels and wavelet coefficients are represented in Mallat form
2. Coefficients are quantized by using nonlinear estimation
3. Lower frequency sub-band is transmitted/stored
4. Highest frequency sub-band of each components are quantized with zero-tree wavelet coefficient concept introduced in this scheme
5. Each, nonzero coefficients of highest frequency sub-bands are \( (2 \times 2) \) are coded with quad-tree address and transmitted/stored
   - 5.1 scale invariant coefficients are coded with their address
   - 5.2 other non-zero coefficients are coded with their address

Next how the weights are determined. Usually weight of the coefficients is half, but some coefficients are scale invariant in this case the weight is one. These weights are determined as follows:

\[
\begin{align*}
\text{if } (c_{i,0} + \ldots + c_{i,3}) < \text{threshold} \\
\quad w_{i,j} = 1 \\
\text{else} \\
\quad w_{i,j} = 1/2
\end{align*}
\]

**Decoding Algorithm:**

1. Lower frequency component coefficients are placed in corresponding array position
2. Highest frequency sub-bands coefficients of scale invariant are placed in the corresponding position of the array
3. Next high frequency sub-bands are estimated recursively: each coefficients having weight one; corresponding coefficient is the summation of initial distribution (coefficients)
4. Next high frequency sub-bands (other coefficients that are not estimated from the previous step) are estimated recursively: each coefficients having weight 1/2; corresponding coefficient is the half of the summation of initial distribution (coefficients)
5. Inverse wavelet transform is obtained to reconstruct the image with the array
In order to evaluate the results root mean squared (RMS) error and peak-to-peak signal-to-noise ratio (PSNR) values are used against the number of nonzero wavelet coefficients. An image Lena $512 \times 512$ is used as the test image. The high-frequency nonzero wavelet coefficients are coded with SWHFC and the quantized coefficients from the same scale are used to reconstruct the image as with linear approximation. The results are compared with other methods in the table 1.

The other methods such as SPIHT, FZW etc., give slightly better results than SWHFC, but advantages of this method remains as with the WFA image coding over the fractal image coding. Also, it describe an novel algorithm to encode an image wavelet coefficients with the WFA coding efficiently. This is a significant improvement in coding wavelet coefficients with WFA.

It is obvious that SWHFC gives high compression at a given fidelity. The coded coefficients with this method may be further coded with an entropy coder to obtain further compression. Fidelity increase with increasing the non zero wavelet coefficients and/or progressive fidelity transmission.

Haar orthonormal basis function is chosen as the wavelet transform, since the images with edges give high-frequency coefficients than the smooth wavelets, as this method depends on the self-similarity between the high-frequency components. With the increasing threshold value or reducing the number of coefficients, increase in tiling effect. Also, WFA coding of pixels introduces blocking artifact, but this effect is greatly reduced as the wavelet coefficients are coded. This is the obvious advantage in using wavelet coefficients with WFA rather than pixel coding with WFA. It is a better approach than the linear approximation.

Table 1: Comparison of algorithms for Lena image $512 \times 512$

<table>
<thead>
<tr>
<th>Method</th>
<th>CR</th>
<th>RMSE</th>
<th>PSNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SWHFC</td>
<td>0.036bpp</td>
<td>14.05</td>
<td>25.17</td>
</tr>
<tr>
<td>SQS [12]</td>
<td>0.036bpp</td>
<td>-</td>
<td>25.86</td>
</tr>
<tr>
<td>WFC [13]</td>
<td>0.036bpp</td>
<td>-</td>
<td>26.42</td>
</tr>
<tr>
<td>SPIHT [14]</td>
<td>0.036bpp</td>
<td>-</td>
<td>26.49</td>
</tr>
<tr>
<td>FZW [11]</td>
<td>0.036bpp</td>
<td>-</td>
<td>26.49</td>
</tr>
<tr>
<td>Spatial fractals [7]</td>
<td>17.4</td>
<td>-</td>
<td>24.90</td>
</tr>
<tr>
<td>DCT fractals [10]</td>
<td>18.5</td>
<td>-</td>
<td>26.10</td>
</tr>
<tr>
<td>JPEG</td>
<td>0.15bpp</td>
<td>-</td>
<td>26.44</td>
</tr>
</tbody>
</table>

8. Conclusion

References


[3] U.Hafner, \textit{Asymmetric Coding in (m)-WFA Image Compression}, preprint


Figure 3: Lena original image $512 \times 512$ and Quantized Haar coefficients.

Figure 4: Estimated coefficients and Reconstructed image 25.17dB at 0.03bpp $512 \times 512$. 