Improved SPIHT Algorithm

Humberto de Jesús Ochoa Domínguez, Vianey Guadalupe Cruz Sánchez
Dept. of Electrical and Computer Engineering
Universidad Autónoma de Ciudad Juárez
Ciudad Juárez, Chihuahua, México
{hochoa, vianey.cruz}@uacj.mx

Osslan Osiris Vergara Villegas
Dept. of Industrial and Manufacturing Engineering
Universidad Autónoma de Ciudad Juárez
Ciudad Juárez, Chihuahua, México
overgara@uacj.mx

Abstract— An improved SPIHT algorithm that combines the SPIHT and the subbands threshold calculation to reduce the number of comparison operations, without sacrificing the reconstructed image quality, is proposed. After applying discrete wavelet transform to the image, the threshold of each independent subband is calculated. The scanning of the sets inside a subband is determined by the magnitude of the thresholds that establish a hierarchical scanning not only for the set of coefficients with large magnitude, but also, for the subbands. The proposed algorithm uses the set partitioning technique to sort the transformed coefficients. Results show that the proposed algorithm significantly reduces the number of comparison operations in the sorting passes while maintaining the visual quality and PSNR of the recovered image.

Keywords- Image processing, image coding, SPIHT, transform coding, embedded coding.

I. INTRODUCTION

It has been proved that image compression algorithms based on the Discrete Wavelet Transform (DWT) provide high coding efficiency for natural images [1 – 6]. Besides, DWT has desirable properties such as efficient multisresolution representation, scalability and embedded coding with progressive transmission, which are suitable for image compression.

The Set Partitioning In Hierarchical Trees (SPIHT) encoder can achieve rate scalability and high coding performance [6 – 11]. The encoder makes use of the progressive transmission to exploits the self-similarity of the DWT coefficients across different scales. Also, in order to iteratively extract the most important information in each subband, SPIHT uses partial ordering by comparing the transformed coefficient's magnitudes with a set of octave decreasing thresholds.

Even when SPIHT is not a standard or a widely used scheme in user applications, it has become a well-known algorithm in the image coding community. There are many video applications that use image compression with the embedded coding property [17 – 19] and medical imaging applications with SPIHT as a core [20].

One of the main drawbacks of the SPIHT is the slow coding speed because of the dynamic processing order that depends on the image contents. In the past, several works to overcome the shortcoming have been proposed [10], [12], [13]. The efforts to improve the performance have been mostly concentrated in modifying the lists used to store the sets coordinates.

In this paper, we propose an improved SPIHT to reduce the number of comparison operations, while keeping the PSNR and the visual quality of the recovered images. The independent thresholds of each subband are calculated and the scanning sequence of the sets is modified according to the magnitude of the current threshold. Therefore, sets inside subbands with a high threshold are scanned first. Furthermore, SPIHT calculates only one maximum threshold, which is quantized after one pass. The only information available is what subband contains the highest threshold. There is no information about the magnitude of sets across scales. Sets of a tree are scanned from father to children to grand children when is needed, even though the entire tree is not significant to a current threshold adding operations. However, in the proposed algorithm, each subband threshold is known and if any is lower than the current threshold, the corresponding subband sets are not scanned, saving comparison operations. Besides, list of insignificant sets (LIS) grows adaptively, according to the bit rate required.

II. SPIHT ANALYSIS

The transformed image $\chi$ has a hierarchical pyramidal structure as shown in figure 1. The highest frequency subband has the finest samples and lies at the bottom right of the pyramid. The lowest frequency subband has the coarsest samples and lies at the top left of the pyramid. The last scale of decomposition yields four subbands including the lowest frequency subband. Therefore, if the image is decomposed using $L$ scales, a total of $3L+1$ subbands are obtained. At each scale $l$, the three high frequency subbands capture the horizontal, vertical, and diagonal directions.

The maximum threshold ($T_{max}$) is calculated based on the number of bits needed to represent the highest transform coefficient in $\chi$. $T_{max}$ becomes the current threshold ($T_c$) for the first pass which is sent to the decoder, followed by the $T_c^{th}$ most significant bit of the coefficients and their correspondent sign. This is called the sorting or dominant pass. In a refinement pass, the $T_c^{th}$ most significant bit, from the previous significant samples, is sent to the decoder in the same order as its coordinates. The sorting passes also divide the set of coefficients into partitioning subsets of samples. Each set is compared against the threshold. If the coefficients inside a set are insignificant (magnitude lower than $2^{T_c}$) a
‘0’ is sent to the decoder. The set is scanned again in subsequent sorting passes until any of its coefficients results significant to $T_c$. Then, the set is partitioned into new four subsets. The division continues until the test is done to all single coordinate significant subset in order to identify each significant coefficient. Afterward, the threshold is decreased by one and taken as the new reference or current threshold ($T_c$) for the next pass.

SPIHT uses spatial orientation trees to organize sets. Each node of the tree can have four descendants or offspring (sets of 2x2 adjacent samples) of the same spatial orientation in the next finer level of the pyramid, or no offspring. In addition, SPIHT maintains three different lists: one list of insignificant sets (LIS), one list of significant pixels or samples (LSP), and one list of insignificant pixels (LIP). LSP contains the coordinates of significant coefficients of a specific threshold. These coefficients are represented in the bit stream with the most significant bit (MSB) and the sign bit. LIP contains the coordinates of coefficients that are not significant to the current or previous thresholds. LIP may contain coordinates of samples of sets with at least one significant sample. Coefficients in LIP are represented in the bit stream with one bit, usually ‘0’, which is the MSB of the insignificant coefficient. LIS contains the coordinates of sets, which have not been significant to either the current or previous thresholds.

Suppose that the highest threshold ($T_c = T_{max}$) of an image of size $M \times N$, transformed into the wavelet domain using $L$ levels of decomposition, is represented with $B$ bits and that the threshold is in the lowest resolution, lowest frequency subband ($LL_L$). Furthermore, suppose that the transformed image is going to be coded using the SPIHT coder and the threshold of each subband to have knowledge of the highest coefficients magnitude per subband avoiding scanning subbands with thresholds lower than $T_c$. Hence, the encoder has to calculate the thresholds of each subband to have knowledge of the highest coefficients magnitude per subband avoiding scanning subbands with thresholds lower than $T_c$.

The contribution is only from subbands with a threshold higher or equal to the current threshold. Hence, the encoder has to calculate the thresholds of each subband to have knowledge of the highest coefficients magnitude per subband avoiding scanning subbands with thresholds lower than $T_c$. In other words, the thresholds help avoid unnecessary entries to LIS, that contains sets whose coefficients have not been significant in previous passes. These entries add more complexity to the process.

III. PROPOSED METHOD

The algorithm is applied to a hierarchical pyramidal structure obtained by using the DWT on the image. Figure 1 shows a two level of DWT decomposition. The rightmost pyramid shows the depth of the coefficients; each block represents one bit. The topmost block is the most significant bit.

As in SPIHT, the algorithm defines 4 steps: 1) initialization, 2) sorting (dominant) pass, 3) refinement pass and 4) quantization step. Unlike the SPIHT, during the initialization step, each subband threshold is calculated using equation (4) and stored in a short register of length $3L+1$. The current threshold is initialized to the maximum threshold ($T_c = T_{max}$).

$$I_{l,k} = \begin{cases} 1 & T_c > T_{h_{l,k}} \\ 0 & \text{Otherwise} \end{cases}$$

$$I_{l,k}' = \begin{cases} 1 & T_c \leq T_{h_{l,k}} \\ 0 & \text{Otherwise} \end{cases}$$

The contribution is only from subbands with a threshold higher or equal to the current threshold. Hence, the encoder has to calculate the thresholds of each subband to have knowledge of the highest coefficients magnitude per subband avoiding scanning subbands with thresholds lower than $T_c$. In other words, the thresholds help avoid unnecessary entries to LIS, that contains sets whose coefficients have not been significant in previous passes. These entries add more complexity to the process.

$$C = M \times N \sum_{l=0}^{L} \sum_{k=1}^{2^l} \frac{1}{2^{l-2}} \sum_{i=1}^{2^l - 1} B_{i,k}$$

where

$$Th_{l,k} = \log_2 \left( \max_{i,j \in \mathbb{X}} |s_{i,j}(i,j)| \right)$$

Figure 1. Two levels of hierarchical pyramidal decomposition of an image, $L=2$. 

As in SPIHT, the algorithm defines 4 steps: 1) initialization, 2) sorting (dominant) pass, 3) refinement pass and 4) quantization step. Unlike the SPIHT, during the initialization step, each subband threshold is calculated using equation (4) and stored in a short register of length $3L+1$. The current threshold is initialized to the maximum threshold ($T_c = T_{max}$).
where \( s_{l,k} \) is the \( k \)th subband at scale \( l \), \( cs_{l,k} \) is a coefficient inside the subband \( s_{l,k} \) and \( \lfloor \cdot \rfloor \) is a floor operation. LSP and LIP are initialized as in SPIHT. In the sorting pass, significant coefficients are compared in LIP using (5).

\[
2^{T_{c+1}} > \max_{[i,j] \in X} \left\{ |cs_{l,k}[i,j]| \right\} \geq 2^{T_c}
\]

(5)

If the coefficient is significant a 1 is sent to the decoder followed by its sign bit. Afterwards, the position is removed from the LIP and set in the LSP. If the coefficient is not significant a 0 is sent to the decoder leaving the LIP unchanged. The significance of a set of coordinates \( T \) is tested using (6).

\[
S_{l,k}(T) = \begin{cases} 
1 & \text{max}_{[i,j] \in T} |cs_{l,k}[i,j]| \geq 2^{T_h} \\
0 & \text{otherwise.}
\end{cases}
\]

(6)

The register of thresholds is inspected and if a threshold is equal to the new \( T_c \) its corresponding subband sets are added to the LIS. As we can see, the encoder saves operations, because the sets of a subband are added to the LIS adaptively according to the transformed characteristics of the image as in (3), saving entries to this list. Then, the LIS is analyzed; a ‘1’ is sent if a set is insignificant to \( T_c \) or ‘0’ otherwise. After that, the set and significant coefficients inside of it are treated as in SPIHT. In the refinement pass, each element in the LSP, resulting in previous passes, is compared with \( T_c \). The resulting bit is sent to the decoder. In the quantization step \( T_c \) is decremented by 1, to obtain a new \( T_c \), and a new sorting pass is started. The bit rate is controlled using a counter whenever a bit is added to the bit stream. The total bit stream is computed by,

\[
\text{Total bitstream} = M \times N \times \text{Bit rate}
\]

(7)

IV. RESULTS

The well-known images of Lena, Barbara and Mandrill, 8 bpp, gray scale and 512 x 512 size were used to test the algorithm [14]. The energy compaction is one of the most important metrics in the evaluation of filters used in the transform coding schemes. It refers to the property on how the energy in the transform domain is distributed among its coefficients. For encoding purposes, it is required that most of the energy of an image is concentrated in a few coefficients. In this work, the 9-7 tap biorthogonal wavelet filter bank - Cohen-Daubechies-Feauveau (CDF 9/7) - is used to obtain the pyramidal decomposition because it is better than others in terms of energy compaction [1], [21], with reflection extension to each image and \( L = 6 \). No entropy coder was used on the resulting bitstream. Comparison operations, peak-signal-to-noise-ratio (PSNR) and visual quality performance were compared with SPIHT [15].

Figure 2 shows the number of comparison operations for the (a) Lena, (b) Barbara, (c) Mandrill images of the proposed method compared with SPIHT and (d) the number of operations saving. Because of the embedded
characteristics of the code, operations in previous passes are added to the number of operations in the current pass making SPIHT to start from a high number of operations even for very low bit rates because of (2).

SPIHT needs about 9.5 millions of comparison operations in sorting passes to reach a bit rate of 0.0625 bpp, while the proposed method invest only 0.65, 1.3 and 1.67 millions respectively for the same images; to reach 4 bpp in Lena and Barbara images, SPIHT accumulates about 28.4 millions of comparison operations and 24.5 in Mandrill, while the proposed method requires about 13.8, 15.8 and 13.7 millions of comparison operations.

B. PSRN Measure

For images represented by 8 bits per pixel the PSNR in dBs is defined by,

$$PSNR = 10 \log_{10} \frac{255^2}{\frac{1}{MN} \sum_{i,j} (x(i,j) - \hat{x}(i,j))^2}$$  \hspace{1cm} (8)

where $x(i,j)$ and $\hat{x}(i,j)$ denote the original and reconstructed image values, respectively, at coordinates $[i,j]$. Table I shows the PSNRs at different bit rates (BR).

### Table I. PSNR in dB for the Lena, Barbara and Mandrill Images.

<table>
<thead>
<tr>
<th>BR (bpp)</th>
<th>Lena</th>
<th>Barbara</th>
<th>Mandrill</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>22.46</td>
<td>22.54</td>
<td>20.09</td>
</tr>
<tr>
<td>0.0625</td>
<td>27.81</td>
<td>27.87</td>
<td>23.25</td>
</tr>
<tr>
<td>0.10</td>
<td>29.82</td>
<td>29.81</td>
<td>24.27</td>
</tr>
<tr>
<td>0.25</td>
<td>33.65</td>
<td>33.65</td>
<td>27.42</td>
</tr>
<tr>
<td>0.50</td>
<td>36.78</td>
<td>36.78</td>
<td>31.23</td>
</tr>
<tr>
<td>0.75</td>
<td>38.34</td>
<td>38.36</td>
<td>33.92</td>
</tr>
<tr>
<td>1.00</td>
<td>39.88</td>
<td>39.88</td>
<td>36.05</td>
</tr>
<tr>
<td>2.00</td>
<td>44.11</td>
<td>44.05</td>
<td>41.90</td>
</tr>
<tr>
<td>3.00</td>
<td>47.96</td>
<td>47.87</td>
<td>46.21</td>
</tr>
<tr>
<td>4.00</td>
<td>52.30</td>
<td>52.21</td>
<td>50.40</td>
</tr>
</tbody>
</table>

Note that the proposed method and SPIHT follow close to each other, except for high bit rates where the proposed method performs slightly better because of the bit saving to signal the significance of sets.

C. Structural Similarity Index Measure (SSIM)

The visual quality assessment of the recovered images was addressed according to the structural similarity index measure (SSIM) [16]. The closer the index to 1, the more similar the two compared signals are. The recovered images were compared with the original one.

The closer the indexes to 1, the more similar the two compared signals are. Recovered images were compared with the original one. Table II shows the results of the modified and the SPIHT methods. Again, both algorithms perform similar to each other except for Barbara and Mandrill that for high bit rates (above 0.5 bpp) the modified SPIHT performs slightly better.

### Table II. Structural Similarity Index Measure (SSIM) for the Lena, Barbara and Mandrill Images.

<table>
<thead>
<tr>
<th>BR (bpp)</th>
<th>Lena</th>
<th>Barbara</th>
<th>Mandrill</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>0.626</td>
<td>0.631</td>
<td>0.515</td>
</tr>
<tr>
<td>0.10</td>
<td>0.893</td>
<td>0.892</td>
<td>0.797</td>
</tr>
<tr>
<td>0.25</td>
<td>0.952</td>
<td>0.951</td>
<td>0.894</td>
</tr>
<tr>
<td>0.50</td>
<td>0.974</td>
<td>0.975</td>
<td>0.957</td>
</tr>
<tr>
<td>1.00</td>
<td>0.987</td>
<td>0.988</td>
<td>0.982</td>
</tr>
</tbody>
</table>

V. CONCLUSIONS

This paper presented an improved SPIHT algorithm that combines the advantages of the SPIHT and the hierarchical sorting of the subbands to reduce the number of comparison operations during the sorting pass and, at the same time, to limit the number of entries to LIS. Energy compaction and ordering are two essential requirements for a good compression. The DWT CDF 9/7 fulfilled the first requirement and the second requirement was improved by scanning the subbands sets according to their corresponding hierarchy imposed by the thresholds. The scanning fashion helps drastically reduce the number of entries to LIS, making the list to grow according to the
bit rate and reducing the number of comparison operations needed in a sorting pass. The proposed algorithm starts from a reduced number of comparison operations and also ensures that higher magnitude coefficients are found soon. The proposed system keeps the same visual quality and PSNR as the SPIHT.

REFERENCES


