RC 21010 (Oct. 22, 1997)
Computer Science/Mathematics

IBM Research Report

Entropy Coding Techniques for Lossless Image Compression with Reversible Integer Wavelet Transforms

Nasir Memon
Computer Science Department
Northern Illinois Univ.
DeKalb, IL 60115
Email: memon@cs.niu.edu

Xiaolin Wu
Computer Science Department
Univ. of Western Ontario
Canada N6A 5B7
Email: wu@csd.uwo.ca

Boon-Lock Yeo
IBM Research Division
T.J. Watson Research Center
Yorktown Heights, NY 10598
Email: yeo@watson.ibm.com

IBM Research Division
Almaden · T.J. Watson · Tokyo · Zurich

ENTROPY CODING TECHNIQUES FOR LOSSLESS IMAGE COMPRESSION WITH REVERSIBLE INTEGER WAVELET TRANSFORMS

[1]Nasir Memon* [2]Xiaolin Wu and [3]Boon-Lock Yeo

[1]Computer Science Department, Northern Illinois Univ., DeKalb, IL 60115. memon@cs.niu.edu
[2]Computer Science Department, Univ. of Western Ontario, Canada N6A 5B7. wu@csd.uwo.ca
[3]IBM T.J. Watson Research Center Yorktown Heights, NY 10598. yeo@watson.ibm.com

ABSTRACT

The past few years have seen an increasing interest in using reversible integer wavelets in image compression. Reversible integer wavelet image coders facilitate decompression from low bit rates all the way up to lossless reconstruction. However, in the past, specific implementations of such techniques, like S+P, could not match the lossless compression performance of state-of-the-art predictive coding techniques like CALIC. We demonstrate for the first time that reversible integer wavelets together with proper context modeling can match the lossless compression performance of CALIC. This can be done without increase in complexity over S+P. Our findings present a strong argument for using subband coding as a unified, elegant approach for both lossy and lossless image compression. Specifically, in this paper we outline how to obtain significantly higher coding efficiency over S+P by utilizing better filters and better context modeling and entropy coding of wavelet coefficients.

1. INTRODUCTION

Among various methods which have been devised for lossless compression, predictive techniques are perhaps the most simple and efficient.

An alternative approach to lossless image compression, that has emerged recently is based on subband decomposition. Subband decomposition provides a way to cope with non-stationarity of image data by separating the information into several scales and exploiting correlations within each scale as well as across scales. A subband approach also provides a better framework for capturing global patterns in the image data. Finally, the wavelet transforms employed for subband decomposition, can be viewed as a prediction scheme (as in [4, 2]), that is not restricted to a casual template but makes a prediction of the current pixel based on 'past' and 'future' pixels with respect to a spatial raster scan.

In addition to the above, there are other advantages offered by a subband approach for lossless image compression. The most important of which is perhaps the natural integration of lossy and lossless compression that becomes possible. By transmitting entropy coded subband coefficients in an appropriate manner, one can produce an embedded bit stream that permits the decoder to extract a lossy reconstruction at a desired bit-rate. This enables progressive decoding of the image that can ultimately lead to lossless reconstruction [4, 7]. The image can also be recovered at different spatial resolutions. These features are of great value for specific applications like teleradiology and the World-Wide-Web, and for applications in "network-centric" computing in general.

The above facts have caused an increasing popularity of the subband approach for lossless image compression. Some excellent work has already been done towards applying subband image coding techniques for lossless image compression, such as S+P [4] and CREW [7]. Although both these techniques provide the attractive features that are made possible by a subband approach, they suffer from poor compression performance in the lossless mode.

In Table 1 we list compression results achieved with the luminance (Y) planes of the ISO test images, each at a spatial resolution of 576×720. We show bit rates achieved with two currently known predictive techniques. The first, CALIC [6], is a context-based predictive technique of moderate complexity, that to the best of out knowledge, gives superior compression performance on the average over currently known practical and general purpose lossless compression techniques The second technique, JPEG-LS, is the recently finalized international standard for lossless and near-lossless image compression [1]. JPEG-LS is a low complexity lossless image coding technique that, given its simplicity, provides compression performance within a few percent of the best known techniques. Finally, in the third and fourth columns of Table 1 we also show bit rates achieved with the arithmetic coding and Huffman coding versions of the multi-resolution lossless image coding technique S+P (S-Transform with Prediction) [4]. For the S+P schemes, four levels of wavelet decomposition were performed.

On examining the results in Table 1 we see that the compression performance obtained by the arithmetic coding version of S+P is significantly lower than CALIC. Similarly, the Huffman coding version of S+P does not do better than JPEG-LS that uses Golomb-Rice codes which are essentially Huffman codes.

In this paper we diminish the performance gap between CALIC and subband techniques. We do so by using better integer reversible wavelets and some recently developed modeling and entropy coding techniques for lossless image

*N. Memon was supported by NSF Career award NCR 9703969

Image	CALIC	JPEG-LS	S+P (Arith)	S+P (Huff)
Balloon	2.78	2.90	2.97	3.14
Barb 1	4.33	4.69	4.53	4.93
Zelda	3.72	3.89	3.84	3.98
Hotel	4.22	4.38	4.53	4.83
Barb 2	4.49	4.69	4.71	5.08
Board	3.50	3.68	3.82	4.07
Girl	3.71	3.93	3.96	4.15
Gold	4.38	4.48	4.56	4.81
Boats	3.77	3.93	4.03	4.34
Average	3.88	4.06	4.11	4.37

Table 1: Compression results in bits per pixel (bpp) for three popular, state-of-the-art lossless coders, CALIC, JPEG-LS, and S+P.

compression [3].

2. FILTER DESIGN AND SELECTION.

In recent work, Calderbank et al. [2] have shown how to obtain wavelet transforms that map integers to integers. They have also demonstrated that better lossless coding efficiency can be obtained simply by using "better" filters. In particular, the reversible integer version of the (4,2) and (2+2,2) filters [2] have the best overall performance and outperform the filters of S+P [4] and CREW [7].

We denote by $(s_{0,j})_j$ the original signal of interest, $(s_{1,j})_j$ and $(d_{1,j})_j$ the lowpass and highpass coefficients respectively after a wavelet transform. The filters described in [4, 7] takes the following form:

$$
\begin{aligned}
d_{1,l}^{(1)} &= s_{0,2l+1} - s_{0,2l} \\
s_{1,l} &= s_{0,2l} + \lfloor d_{1,l}^{(1)}/2 \rfloor \\
d_{1,l} &= d_{1,l}^{(1)} + \lfloor \alpha_{-1}\,(s_{1,l-2} - s_{1,l-1}) + \alpha_0\,(s_{1,l-1} - s_{1,l}) \\
&\quad + \alpha_1\,(s_{1,l} - s_{1,l+1}) - \beta_1\, d_{1,l+1}^{(1)} \rfloor .
\end{aligned}
$$

For the S+P transform [4], $\alpha_{-1} = 0, \alpha_0 = 2/8, \alpha_1 = 3/8$ and $\beta_1 = -2/8$. The TS-transform proposed in [7] is another special case, when $\alpha_{-1} = \beta_1 = 0$ and $\alpha_0 = \alpha_1 = 1/4$.

The (4,2) and (2+2,2) implemented as reversible integer wavelet transforms are respectively

$$
\begin{aligned}
d_{1,l} &= s_{0,2l+1} - \lfloor 9/16\,(s_{0,2l} + s_{0,2l+2}) - \\
&\qquad 1/16\,(s_{0,2l-2} + s_{0,2l+4}) + 1/2 \rfloor \\
s_{1,l} &= s_{0,2l} + \lfloor 1/4\,(d_{1,l-1} + d_{1,l}) + 1/2 \rfloor ,
\end{aligned}
$$

and

$$
\begin{aligned}
d_{1,l}^{(1)} &= s_{0,2l+1} - \lfloor 1/2(s_{0,2l} + s_{0,2l+2}) + 1/2 \rfloor \\
s_{1,l} &= s_{0,2l} + \lfloor 1/4(d_{1,l-1}^{(1)} + d_{1,l}^{(1)}) + 1/2 \rfloor \\
d_{1,l} &= d_{1,l}^{(1)} - \lfloor 1/16\,(s_{1,l} - s_{1,l-1} + s_{1,l+1} - s_{1,l+2}) + 1/2 \rfloor .
\end{aligned}
$$

Both the (4,2) and (2+2,2) filters have 4 vanishing moments of the analyzing high pass filter, and 2 vanishing moments of the synthesizing high pass filter. We can interpret the $d_{1,l}$ expression of (4,2) and $d_{1,l}^{(1)}$ expression of (2+2,2) as non-causal predictions of the odd sample $s_{0,2l+1}$ from the neighboring even samples. The update step in evaluating $s_{1,l}$ provides a 'smoother' lowpass signal as compared to direct down-sampling, i.e., using $s_{1,l} = s_{0,2l}$. The extra step in evaluating $d_{1,l}$ of (2+2,2) increases the number of vanishing moments of the analyzing high pass from 2 to 4. To compare their performance against the S+P filter, we use these filters with the S+P's entropy coder, and tabulate the result in Table 2. Results shown in the 3 columns were obtained by exactly the same technique and implementation except for the filter employed. From Table 2 it is clear that just using a different set of filters can lead to significant improvement in performance on some images and also on the average. In the rest of the paper we shall use the (4,2) and the (2+2,2) filters.

Image	(2+2,2)	(4,2)	S+P
Balloon	2.93	2.94	2.97
Barb 1	4.46	4.47	4.53
Zelda	3.77	3.78	3.84
Hotel	4.48	4.49	4.53
Barb 2	4.68	4.68	4.71
Board	3.74	3.75	3.82
Girl	3.90	3.90	3.96
Gold	4.51	4.51	4.56
Boats	3.96	3.97	4.03
Average	4.05	4.05	4.11

Table 2: Bit rates in bpp with S+P's entropy coder when used with the (2+2,2) and (4,2) filters.

3. ENTROPY CODING OF SUBBAND COEFFICIENTS

Once an image has been decomposed, we need to encode the coefficients within each subband losslessly. Visual inspection of the subbands clearly shows they are correlated and the energy in the high frequency subbands is mostly concentrated in the areas where there is edge activity in the original image. For lossy compression, the main challenge here is the design of an effective quantization scheme that takes into account the different characteristics found within each subband and furthermore, the non-stationary nature within each subband. A spatially adaptive quantization scheme is usually employed within each subband followed by some simple entropy coding of the quantized coefficients.

The above approaches work well for applications where a lossy reconstruction of the image is acceptable. However, because of the difference in goals, the optimal way of exploiting correlations for lossy and lossless compression is usually very different. Lossless compression typically requires more sophisticated entropy coding techniques that employ conditional coding as well as context modeling.

In S+P [4], conditional coding of subband coefficients is done by classifying the magnitude of the coefficients into one of a set of ranges called magnitude sets. The magnitude set number is entropy coded and additional bits that determine the exact magnitude within the range and the sign bit are sent uncoded. Entropy coding of the magnitude set of the coefficient at location (i, j) is done by conditioning it to the context k which is computed by: $k = \min(4, \mu_n) + 5 \times \min(4, P(i, j))$ where μ_n is the arithmetic mean of the coefficients representing the West, North, North East and North West neighbors of the coefficient being coded and $P(i, j)$ is the parent coefficient. This leads to 25 contexts in which either static Huffman or adaptive arithmetic coding is done.

The method we employed to form conditioning states for entropy coding of wavelet coefficients is similar to that used in CALIC. Specifically, we condition the current coefficient c on the energy levels of neighboring coefficients that have already been coded in the current band, in the sister band(s), and in the parent band. We also utilize the orientations of different bands because they provide directional structures of edges and textures in the input image. We estimate $|c|$, the energy level of the current coefficient c in LH, HL, and HH bands to be respectively

$$\begin{aligned} \Delta_{LH} &= \delta + \alpha_{lh}|nn| + \beta_{lh}|p_n|; \\ \Delta_{HL} &= \delta + \alpha_{hl}|ww| + \beta_{hl}|p_w|; \\ \Delta_{HH} &= \delta + \alpha_{hh}|c_{lh}| + \beta_{hh}|c_{hl}|; \end{aligned} \tag{1}$$

where

$$\delta = w_1|n| + w_2|w| + w_3|ne| + w_4|nw| + w_5|p| + w_6|p_s| + w_7|p_e|, \tag{2}$$

n, w, ne, nw, nn, ww are directional notations meaning north, west, north-east, north-west, north-north, west-west; and p, p_n, p_w, p_s, p_e are parent coefficient and those to the north, west, south, and east of p in the parent band. Estimates Δ_{LH}, Δ_{HL}, and Δ_{HH} all have δ, called the kernel estimate, as a common part that depends on the immediate four neighbors of c that have been coded in the current band, and the parent coefficient. Since LH, HL, and HH bands are to be coded in raster scans, the neighboring coefficients to the south and east of the current coefficient are future information that is not available for context modeling. However, coefficients p_s and p_e in the parent band do contain some future information, albeit at a coarser scale, of the source, and they benefit context modeling. Therefore, we also include p_s and p_e in δ. Similar ideas of using future information in context modeling were also proposed in the framework of predictive coding such as the subsampling scheme in [5]. But multiresolution wavelet decomposition retains data coherence much better than simple subsampling.

In addition to the common part δ, Δ_{LH} has two different terms nn and p_n because the LH band has predominantly vertical structures; similarly Δ_{HL} has ww and p_w due to the existence of horizontal structures in the HL band. In the HH band, there usually lack predominant structures in any orientation. Instead Δ_{HH} uses the sister coefficients c_{lh} and c_{hl} that are at the same spatial location. This of course implies that we will code HH band after LH and HL bands to establish the necessary causal relationship.

The weights w_i, $1 \le i \le 7$, α's and β's in estimates Δ's can be determined via linear regression so that Δ's are least-square estimates of $|c|$. This can be done using a training set and even on an image-by-image basis. The empirical results reported by this paper were obtained by using heuristically chosen, image-independent weights. Further statistical analysis of wavelet coefficients is under way. We speculate that further compression gains are possible by using statistically optimized weights.

Image	(2+2,2)		(4,2)	
	Proposed	S+P	Proposed	S+P
Balloon	103629	104906	105759	106343
Barb 1	161869	163158	163296	164662
Zelda	141458	142576	142775	143940
Hotel	164805	165635	166419	167111
Barb 2	173608	174936	174997	176201
Board	136690	138187	138505	139829
Girl	137410	138936	138801	140149
Gold	164937	166838	166384	168185
Boats	143865	145290	145696	146804

Table 3: Number of bytes needed to encode the LH, HL and HH bands after one level of decomposition with proposed conditional coding as opposed to S+P's entropy coding. Arithmetic coding used in all cases.

Conditional coding of coefficients based on Δ's leads to their separation into classes of different variances. Clearly, the conditional entropy $H(c|\Delta)$ can be much lower than the self-entropy $H(c)$ if one does a reasonable job on estimates Δ's. For time and space efficiency, and more importantly to reduce model cost, Δ has to be quantized to a small number of (L) levels. In practice, $L = 8$ to 16 is found to be sufficient. Larger L only improves coding efficiency marginally and beyond a point it begins to hurt due to high model cost. Although the Δ quantizer $Q(\Delta)$ can be optimized off-line by standard dynamic programming techniques in order to minimize the conditional entropy over a training set of images [5], in practice, we used an image-independent Δ quantizer with 16 bins.

In Table 3 we give simulation results on the test set of images with the proposed entropy coding scheme as opposed to those obtained using the S+P's entropy coding technique. In both cases we show results for the three high pass bands LH, HL and HH. The coefficients in these bands comprise 75% of the total coefficients to be encoded.

4. BIAS CANCELLATION

To control model cost, we quantize Δ to a small number of conditioning states, and drive an entropy coder with estimated $p(c|Q(\Delta))$. But if we estimate only the conditional expectation $E\{c|S\}$ in a conditioning state S rather than $p(c|S)$, then we can use a far greater number of conditioning states without running into context dilution problem. These estimates of conditional expectations are then used to center the distribution of c within each S around zero — this process is called bias cancellation. We extend the idea of using a large number of contexts to estimate one or few statistical parameters proposed in [5] to coding of subband coefficients.

In CALIC, contexts are formed by thresholding neighboring pixels with respect to the predicted value into two levels to obtain a binary texture pattern. Other local information is then attached to this to obtain the final context.

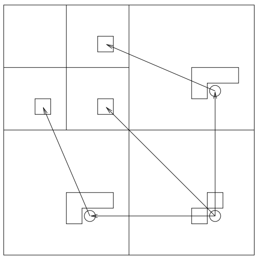

Figure 1: Contexts used for bias cancellation.

Image	(2+2,2)		(4,2)	
	Bias	No Bias	Bias	No Bias
Balloon	102679	103629	104885	105759
Barb 1	159581	161869	161004	163296
Zelda	140569	141458	141913	142775
Hotel	162230	164805	163874	166419
Barb 2	170846	173608	172289	174997
Board	134913	136690	136752	138505
Girl	135968	137410	137359	138801
Gold	163470	164937	164906	166384
Boats	141482	143865	143346	145696

Table 4: Number of bytes needed to encode the LH, HL and HH bands after one level of decomposition, with and without bias cancellation.

This is a reasonable approach given the smoothness of the intensity function in a typical gray level image. However, the high pass bands of a subband decomposed image are not 'smooth'. They are mostly of low magnitude with their sign fluctuating more or less randomly and high magnitudes are found in locations corresponding to edges in the image. Given this fact, we formed the context by using the four immediate neighboring coefficients and the parent coefficient. This is shown in Figure 1. For the HH band, the two corresponding coefficients in the LH and HL band were used instead of the diagonally adjacent neighbors. The context in the HH band is also shown in Figure 1. Coefficients within the context were then quantized into five levels using ± 3 and ± 13 as thresholds, yielding $5^4 = 625$ contexts. Bias cancellation was then performed based on the accumulated sample means within each context. Table 4 gives results for encoding the high pass bands, LH, HL and HH after one level of decomposition. Results are given with and without bias cancellation, with the (2+2,2) and (4,2) filters. The bias canceled coefficients were encoded by technique described in the previous section. It is seen that bias cancellation always results in a decrease in bit rates though often this decrease is small.

5. ENCODING THE LL BAND

In order to encode the LL band it can be decomposed again and similar techniques outlined above can be used to encode the high pass bands at this second level of decomposition. This procedure can be carried on until the low pass band is of very small size and can be transmitted in an uncoded manner, as is done in S+P. However, we have observed that

Image	(2+2,2)	(4,2)	S+P	CALIC	JPEG-LS
Balloon	2.78	2.82	2.97	2.78	2.90
Barb 1	4.34	4.34	4.53	4.33	4.53
Zelda	3.70	3.70	3.84	3.72	3.84
Hotel	4.30	4.30	4.53	4.22	4.53
Barb 2	4.55	4.55	4.71	4.49	4.71
Board	3.53	3.55	3.82	3.50	3.82
Girl	3.74	3.75	3.96	3.71	3.96
Gold	4.37	4.38	4.56	4.38	4.56
Boats	3.79	3.81	4.03	3.77	4.03
Average	3.90	3.91	4.11	3.88	4.06

Table 5: Bits per pixel for test set with (2+2,2) and (4,2) filters and using proposed techniques on LH, HL and HH bands and CALIC on LL band. Results using CALIC, JPEG-LS and full decomposition using S+P are also shown.

compression performance deteriorates with further decomposition. The low pass image is generally very smooth and ideally suited for a predictive coding technique like CALIC or JPEG-LS. Our experimentation indicated that the best overall compression performance is achieved by using one of these techniques for encoding the LL band. Note that by doing this we lose the fully hierarchical nature of techniques like S+P. However, for many applications requiring lossless reconstruction, one level of decomposition may be sufficient.

In table 5 we show the overall bit rates achieved by the proposed techniques applied together, along with CALIC for encoding the LL band. On comparing with the results in Table 1 we see that we have made up almost all of the performance difference between S+P and CALIC.

6. REFERENCES

[1] ISO/IEC JTC 1/SC 29/WG 1. JPEG LS image coding system. ISO Working Document ISO/IEC JTC1/SC29/WG1 N399 - WD14495, July 1996.

[2] A. R. Calderbank, I. Daubechies, W. Sweldens, and B. L. Yeo. Wavelet transforms that map integers to integers. Tech. Rep. , Dept. of Math, Princeton Univ., Sep. 96. To appear in Journal of Applied and Computational Harmonic Analysis.

[3] N. D. Memon and X. Wu. Recent developments in lossless image compression. The Computer Journal, To appear, Sep. 1997.

[4] A. Said and W. A. Pearlman. An image multiresolution representation for lossless and lossy compression. IEEE Trans. on Image Proc., 5(9):1303–1310, 1996.

[5] X. Wu. Efficient and effective lossless compression of continuous-tone images via context selection and quantization. IEEE Trans. on Image Proc., 6(5):656–664, 1996.

[6] X. Wu and N.D. Memon. Context-based adaptive lossless image coding. IEEE Trans. Comm., 45(4):437–444, Apr 97.

[7] A. Zandi, et.al., CREW: Compression by reversible embedded wavelets. In Proc. of DCC, 212–221. IEEE Press, 1995.