

Scalable Lossless and Lossy Image Coding based on the RWHaT+P Pyramid and the Inter-Coefficient Classification Method

Olivier Déforges, Marie Babel, Laurent Bédat, Véronique Coat

▶ **To cite this version:**

Olivier Déforges, Marie Babel, Laurent Bédat, Véronique Coat. Scalable Lossless and Lossy Image Coding based on the RWHaT+P Pyramid and the Inter-Coefficient Classification Method. IEEE International Conference on Multimedia and Expo - ICME, Jun 2008, Germany. pp.185-188. hal-00320852

HAL Id: hal-00320852
https://hal.science/hal-00320852

Submitted on 11 Sep 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

SCALABLE LOSSLESS AND LOSSY IMAGE CODING BASED ON THE RWHAT+P PYRAMID AND THE INTER-COEFFICIENT CLASSIFICATION METHOD

Olivier Déforges, Marie Babel, Laurent Bédat, Véronique Coat

IETR UMR CNRS 6164, Image and Remote Sensing Group, FRANCE
INSA of Rennes, 35043 RENNES. Contact: {odeforge, mbabel,lbedat, coat}@insa-rennes.fr

ABSTRACT

Next generations of still image codecs should not only have to be efficient in terms of compression ratio, but also propose other functionalities such as scalability, lossy and lossless abilities, region of interest coding, etc. In previous works, we have proposed the LAR compression method covering these requirements. In particular, the RWHaT + P pyramid has recently been presented as a powerful reversible scalable coding technique. This paper introduces new significant improvements by the use of an inter-coefficient classification method. Results are discussed and compared to the state of the art.

1. INTRODUCTION

Despite many drawbacks and limitations, JPEG is still the most commonly-used compression format in the world. JPEG2000 overcomes this old technique, particularly at low bit rates, but at the expense of a significant complexity overhead. Therefore, the JPEG normalization group has recently proposed a call for proposals on JPEG-AIC (Advanced Image Coding) in order to look for new solutions for still image coding techniques [1]. Its requirements reflect the earlier ideas of Amir Said [2] for a good image coder: compression efficiency, scalability, good quality at low bit rates, flexibility and adaptability, rate and quality control, algorithm unicity (with/without losses), reduced complexity, error robustness (for instance in a wireless transmission context) and region of interest decoding at decoder level. A. Said and the JPEG committee should add additional functionalities such as image processing at region level, both at the coder or the decoder. The LAR (Locally Adaptive Resolution) tries to address all these features. In [3], we proposed an original scheme able to perform efficient lossy compression, enabling an unusual hierarchical region representation (without any shape description). Then, in [4], we presented an extension of a more efficient scalable multi-resolution solution in terms of both lossy and lossless compression. New improvements have been shown by the introduction of the Reversible Walsh Hadamard Transform (RWHaT) in [5]. This paper addresses further advances though the concept of inter-coefficient classification, focusing on the lossless compression context.

Section 2 gives an overview of the LAR coding method and the Reversible Walsh Hadamard Transform. These two concepts lead to the introduction of the RWHaT pyramid which can be improved when adding a prediction step. Both of these features are presented in section 3. Significant further improvements in terms of lossless compression efficiency are introduced in section 4. Finally, section 5 concludes the paper.

2. THE LAR AND RWHAT METHODS

2.1. LAR overview

The basic concept of the LAR method is that local resolution should be adapted to suit local activity. Also assuming that an image consists of global information and local texture, we firstly proposed a two-layer, content-based codec, both relying on a quadtree partition [3]. The first layer, called the FLAT LAR, encodes the global information at block level representation. The additional second layer enables texture compression within blocks using variable block size DCT/Hadamard transforms. Therefore, the method provides natural SNR scalability. The block sizes are estimated through a local morphological gradient. The direct consequence is that the smallest blocks are located round the edges whereas large blocks map homogeneous areas. This being so, the main feature of the FLAT coder consists of preserving contours while smoothing homogeneous parts of the image. This characteristic is also exploited to get a free hierarchical region representation: from the low bit-rate image compressed by the FLAT LAR, both coder and decoder can perform a segmentation process by iteratively merging blocks into regions. A direct application is then Region Of Interest (ROI) enhancement, by first selecting regions and enabling second layer coding only for the relevant blocks.

2.2. The RWhaT Transform

In order to enable spatial scalability and lossless compression, we studied different pyramidal representations. Lossless compression means that the construction and decomposition steps lead to a reversible process, resulting in the exact recovery of each element of the original source data. Moreover, the quadtree partition requires dyadic decomposition. Some integer wavelet kernels fit these requirements. For instance, Lux [6] proposed a modified version of the 1D Walsh Hadamard Transform (WHT), which has been popularized by Said [2] and known as the "S" transform or "Haar integer wavelet transform"[7]. The S transform is currently known as one of the best integer wavelet bases for reversible compression at the present time[8]. We introduced the Interleaved S+P technique [4] as an improved version of S+P by increasing space prediction. Another solution for a dyadic reversible decomposition has been proposed with the RWHaT transform for 2×2 blocks [5]. The conventional 2D $WHT_{2\times2}$ basis is defined as follows:

$$WHT_{2\times2} \text{ Matrix: } = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \quad (1)$$

Let $U_{2\times2}$ be the input block with:

$$U_{2\times2} = \begin{bmatrix} u_0 & u_1 \\ u_2 & u_3 \end{bmatrix}. \quad (2)$$

The block transformed $Z_{2\times2}$ by the 2D WHT is defined by:

$$\begin{aligned} Z_{2\times2} &= WHT_{2\times2}(U_{2\times2}) = W_{2\times2}\, U_{2\times2}\, W_{2\times2} = \begin{bmatrix} z_0 & z_1 \\ z_2 & z_3 \end{bmatrix} \\ &= \frac{1}{2}\begin{bmatrix} u_0+u_1+u_2+u_3 & u_0+u_1-u_2-u_3 \\ u_0-u_1+u_2-u_3 & u_0-u_1-u_2+u_3 \end{bmatrix}. \end{aligned} \quad (3)$$

Clearly, if the sum of u_i values is odd, then normalization by 2 produces an integer value. On the other hand, if the sum is even, the normalization induces a real value, leading to an ambiguous rounding operation. The solution consists of controlling the rounding step so that the inverse transform can distinguish between odd or even sum cases: for an even sum case, the rounding operations are applied such that the sum of the transformed coefficient is also even. A reversible transform then becomes possible, first by performing the inverse rounding operation to recover actual real transformed values, and secondly by carrying out the inverse 2D WHT transform [5].

3. RWHAT + P PYRAMID

Notations: $I(i,j)$ denotes the pixel in an image I with the coordinates (i,j), $I(b^N(i,j))$ the block $b^N(i,j)$ in I including the set of pixels $\{I(N.i, N.j), \ldots I(N.i+N-1, N.j+N-1)\}$.

3.1. The RWHaT pyramid

Let $\{Y_l\}_{l=0}^{L_{max}}$ be the multiresolution representation of an image I of size $N_x \times N_y$, where L_{max} is the top of the pyramid and $l=0$ the full resolution image. As for the conventional $WHT_{2\times2}$ case, we iteratively construct the pyramid gathering four blocks to form a mean block at the upper level:

$$\left| \begin{array}{ll} l=0, & Y_0(i,j) = I(i,j); \\ l>0, & Y_l(i,j) = \left\lfloor \frac{1}{4}\sum_{k=0}^{1}\sum_{m=0}^{1} Y_{l-1}(2x+k, 2y+m) \right\rfloor \end{array} \right. \quad (4)$$

with $0 \le i \le N_x^l$, $0 \le j \le N_y^l$, where $N_x^l = N_x/l$ and $N_y^l = N_y/l$.

The top-down decomposition of the pyramid consists of encoding the WHT transformed block $Z_l(b^2(i,j))$ of each input block $Y_l(b^2(i,j))$. From (3) and (4), it is clear that the DC component of each block is unambiguously reconstructed from the upper level plus an additional bit called $\epsilon_{z_{0_l}}$. This bit is separately encoded from the other coefficients. Z_{l+1} denotes the WHT transformed block of $Y_l(i,j)$ and $\dot{Z}_l(i,j)$ the corresponding block with $\epsilon_{z_{0_l}}$ as DC value. The reconstruction from the previous level and current WHT transform is given by:

$$\begin{aligned} &\tilde{Y}_l(b^2(i,j)) = EXPAND[Y_{l+1}(i,j)) + \tilde{\dot{Y}}_l(b^2(i,j)] \\ &\text{with } \tilde{\dot{Y}}_l(b^2(i,j)) = WHT_{2\times2}^{-1}\left(\dot{Z}_l\left(b^2(i,j)\right)\right). \end{aligned} \quad (5)$$

The $EXPAND$ function only duplicates a node value from the tree into its four sons. At this step, we get a common pyramidal representation and encoding process based on $WHT_{2\times2}$ transform, but with the exception of a possible lossless decomposition ($\tilde{Y} = Y$ in this case). The top level has been encoded for the two methods by a simple DPCM.

The common pyramid representation involving a dyadic decomposition is generally associated with a multilevel quadtree partition $QP^{[2^{L_{max}}\ldots2^l]}$. Level l of the pyramid also specifies the finest resolution. More generally, we consider here a global quadtree partition of the image $QP^{[N_{max}\ldots N_{min}]}$ according to allowed block sizes. The parameter $N_l \in [N_{max}\ldots N_{min}]$ gives the upper limit of block sizes to be decomposed at level l of the pyramid. For instance, a global partition $QP^{[64\ldots2]}$ leads to the encoding of the representation from sizes 64 to 2, while $N_0 = 8$ means that blocks of sizes 8, 4 and 2 will be decomposed at level 0.

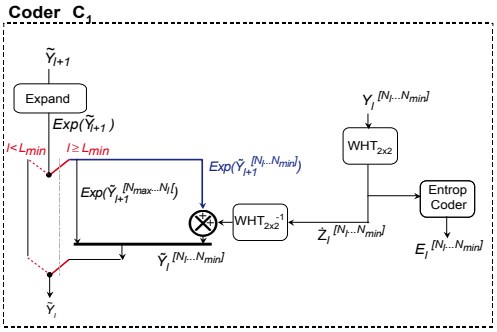


Fig. 1. Simple pyramidal coder: $C1$

Finally, the decomposition parameter L_{min} specifies the last level to be encoded: for all levels lower than L_{min}, the value of all nodes in the pyramid are expanded only. The two pyramid scans presuppose a conditional decomposition. Assuming that $A_l(b^2(i,j))$ is the activity (gradient value) of block $Y_l(b^2(i,j))$, the conditional decomposition during the first pass is defined by:

$$\left| \begin{array}{ll} \text{If } A_l(b^2(i,j)) < Th_{grad} & \text{then encode } Y_l(b^2(i,j)), \\ & \text{else } EXPAND\,[Y_{l-1}(b^2(i,j))] \end{array} \right. \quad (6)$$

with typical $Th_{grad} = 30$. Figure 1 shows the global coding based on this model called C_1 coder, and restricted to the first pass. Blocks that are not decomposed during the first pass are processed during the second one.

The use of quadtree decomposition enables the approach to act as an "objective context modeling", separating the error prediction laws characterized by high entropy for the first pass, and low entropy for the second pass.

3.2. Prediction in the RWHaT pyramid

A prediction step can precede the transform operation to increase compression. Interpolation can also give a suitable means of providing a full resolution image from a decoded level $l > 0$. Even if functions have different objectives, we proposed in [5] a unified estimation solution. The main advantage is that, under particular conditions, the image can be progressively reconstructed from a previously interpolated level which also acts as prediction step. $\check{Y}_l(b^2(i,j))$ and $\tilde{Y}_l(b^2(i,j))$ denote the estimated and reconstructed blocks of $Y_l(b^2(i,j))$ respectively. The estimation

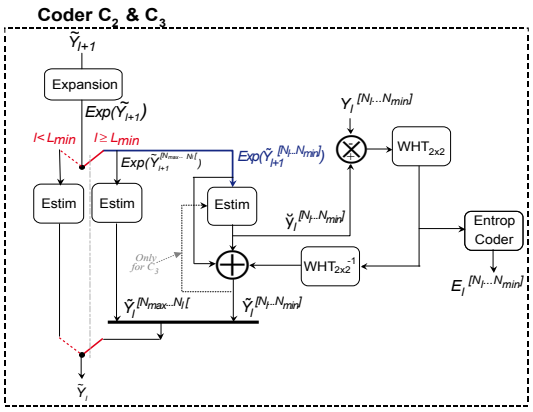


Fig. 2. Pyramidal coder with prediction step: $C2$ and $C3$

process consists of rebuilding unknown values around their block mean in a 2D context, according to a linear function. Two versions have been considered: one exploiting only inter-level information ($C2$), the other one using both inter and intra-level information ($C3$). Figure 2 shows new coding schemes including the estimation stage. The C_2 coder uses only inter-level relationships and can be useful for progressive reconstruction. The C_3 coder also makes use of reconstructed values at the current level and, obviously, leads to better compression.

Both SNR and spatial scalabilities are displayed in figure 3 for a lossless compression process. Six levels of decomposition are considered here, implying eleven successive streams ($1 + 2 \times 5$). At low bit rates, visual distortions are essentially due to a blurring effect and are less disturbing than blocks or ringing artefacts. At the end of the first pass, reconstructed images present perfect accuracy upon strong contours while homogeneous areas appear as smooth regions.

4. INTER-COEFFICIENTS CLASSIFICATION

The *a priori* estimation of entropy laws for error coefficients is a crucial step for the efficiency of a lossless coder. In CALIC [9], which is still one of the best available reversible compression techniques, Wu has introduced the concept of "context modeling" to characterize coefficients from the activity observed in their neighborhood. This method has been widely used in lossless compression context.

Let $\breve{z}_i$ ($0 \leq i \leq 3$) be the transformed coefficients of the estimated block values. $\delta_{z_i} = z_i - \breve{z}_i$ denotes the associated errors. As previously mentioned, the exact reconstruction of the DC value only requires one bit, and is considered separately. Consequently, only $\breve{z}_1$, $\breve{z}_2$ and $\breve{z}_3$ have to be encoded. The context modeling approach suggests the estimation of local energy Δ_i from the estimated value and previous locally-decoded data. To avoid context dilution problems, Δ_i is generally quantized into L levels. The a priori estimation of δ_{z_i} then consists of trying to maximize the conditional probability function $p(\delta_{z_i}|\Delta_i)$. This means that the various quantization levels have to be adapted to each image. In a lossless context and with full resolution processing, the number of encoded values is sufficient for an in-line estimation. However, in a multiresolution scheme, the top levels induce only a few values and they are not sufficient to enable such an approach. Another difficulty occurs when considering a lossy coding context, with quantized values. The problem maximizes the conditional probability $p(\delta_{\hat{z}_i}|\Delta_i)$ (with $\hat{z}_i$ the quantized value of z_i) and no more $p(\delta_{z_i}|\Delta_i)$. $\delta_{\hat{z}_i}$ is, of course, dependent on the quantization factor q_z, and if we consider a linear quantization, we can write

$$p(\delta_{\hat{z}_i}|q_z) = p(q_z\delta_{\hat{z}_i} \leq \delta_{z_i} < (q_z+1)\delta_{\hat{z}_i}). \tag{7}$$

$p(\delta_{\hat{z}_i}|q_z)$ is an increasing function of q_z. As q_z and Δ_i are independent, $p(\delta_{\hat{z}_i}|(\Delta_i, q_z))$ has to be maximized. When Δ_i is quantized on L levels, and q_z on Q_n levels, they define L Q_n contexts, making the estimation process much more complex.

To avoid the context quantization problem, we have introduced the concept of inter-coefficient classification. First, the energy function is defined according to the method features by considering:

1. the dynamics of the estimated value $\breve{z}_i$,
2. the difference δ_{z_0} between the actual mean value of the block, z_0, and the estimated value, $\breve{z}_0$.

The energy function Δ_i associated with δ_{z_i} is defined as follows:

$$\Delta_i = w_0 \ |\delta_{z_0}| + w_1 \ |\breve{z}_i| \,, 1 \leq i \leq 3. \tag{8}$$

The weights w_0 and w_1 have been empirically fixed at 0.5 and 1.

Inter-coefficient classification then consists of dividing the coefficients into equiprobable classes. Fixing the maximal number of classes at Nb_class, the definition of a class $CLASS_n$ is given by:

$$\begin{array}{l} CLASS_n = \left\{\delta_{z_i} | T_{\Delta_{n-1}} > \Delta_i \geq T_{\Delta_n}\right\}, \ 1 \leq n \leq Nb_class \\ \text{with } T_{\Delta_0} = \infty, T_{\Delta_{Nb_class}} = 0 \\ \text{and such as } \forall n \in \{0 \ldots Nb_class\}, \frac{card(CLASS_n)}{Nb_class} = const. \end{array} \tag{9}$$

Each class constitutes a specific substream sent to the entropy coding layer. δ_{z_3} coefficients present lower dynamics than $\delta_{z_{1/2}}$ ones and are then scanned independently.

The $C2$ codec involves only inter-layer prediction, and thereby is slightly less efficient than $C3$. Nevertheless, it provides interesting SNR scalability when used with the classification coefficients method. In particular, as predicted coefficients do not depend on reconstructed data at the same level, values can be independently reconstructed. Thus, only coefficients belonging to $CLASS_0$ for instance can be decoded when providing the class thresholds. This particular class contains coefficients with a higher dynamic, but it is also the most significant in terms of prediction error correction.

The first order entropy for the proposed method ($C3$ mode) is compared in Table 1 with the state of the art: S+P (scalable) and CALIC [9] (non scalable) methods. Both the quadtree partition (Qd) and the inter-coefficient classification (Cl) techniques allow for major improvements. Combining the two techniques leads to an even better solution, significantly outperforming state of the art. This means that the two approaches act differently for data decorrelation and can be complementary.

5. CONCLUSION

The LAR codec is based on the concept of a two-layer encoding process. The first one is used to compress global information, while the second one adds local texture. One scalable and



a) **First pass**, $\mathbf{L_{min} = 5}$ *(1 stream): 0.005 bpp*

b) **First pass**, $\mathbf{L_{min} = 3}$ *(3 streams): 0.071 bpp*

c) **First pass**, $\mathbf{L_{min} = 1}$ *(6 streams): 0.99 bpp*

d) **Second pass**, $\mathbf{L_{min} = 0}$ *(11 streams): 4.84 bpp* (lossless)

Fig. 3. Scalable lossless coding on "Barbara2" with partition $QP^{[64\ldots2]}$

	Entropy (bpp)								
Image	Raw	S	RWHT	CALIC	S+P	RWHaT+P	RWHaT+P Qd	RWHaT+P Cl	RWHaT+P Qd&Cl
Barbara2	7.51	5.45	5.47	4.93	5.04	5.06	4.89	4.90	**4.84**
Hotel	7.57	5.11	5.09	4.57	4.97	4.83	4.60	4.62	**4.54**
Lena	7.44	4.77	4.75	4.33	4.33	4.30	4.19	4.16	**4.14**
Gold	7.60	5.08	5.06	4.65	4.73	4.73	4.66	4.64	**4.61**
Peppers	7.57	4.89	4.87	4.58	4.67	4.54	4.43	4.41	**4.38**
us	4.84	3.65	3.64	3.60	3.78	3.78	3.26	3.25	**3.04**
tools	7.52	5.95	5.95	5.53	5.73	5.71	5.50	5.48	**5.41**
Average	7.15	4.99	4.97	4.60	4.75	4.71	4.50	4.49	**4.42**

Table 1. Comparison of the proposed approaches with state-of-the-art methods S, CALIC and S+P . First-order entropy (bit/pixels).

reversible solution has been proposed with the RWHaT+P pyramid. The two layers of the LAR codec are then implemented as two pyramidal decomposition passes, according to a quadtree partition. In terms of scalability, this process provides spatial and semantic SNR enhancement. In terms of compression, it acts as an objective coefficient classification estimated from the local activity of blocks. We propose here a complementary classification with a finer level of granularity, by considering activity measurements at both block and pixel levels. The inter-coefficient classification introduced in this paper avoids the common problems of context quantization inherent to such approaches, and significantly increases lossless compression efficiency. At the present time, this method also outperforms previous LAR performances for lossy coding. Future work will concentrate on the definition of a suitable entropy layer coder as the present system involves only a simple arithmetic one.

Two other major new features have been also presented, leading to additional decorrelation, and further significant improvements i.e. content-based pyramidal decomposition and inter-coefficient classification. The global scalable coding scheme surpasses both S+P and CALIC. It also provides locally adaptive multiresolution representations for the rebuilding of good quality images at low bit rates.

6. REFERENCES

[1] "Jpeg-aic : scope and evaluation," *International Standards Organization working document, ISO/IEC SC29/WG 1/N4326*, 2007.

[2] A. Said and W. Pearlman, "Reversible image compression via multiresolution representation and predictive coding," in *Visual Communication and Image Processing*. SPIE, Novembre 1993, vol. 209, pp. 664–674.

[3] O. Déforges, M. Babel, L. Bédat, and J. Ronsin, "Color LAR codec: a color image representation and compression scheme based on local resolution adjustment and self-extraction region representation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 8, pp. 974–987, August 2007.

[4] M. Babel, O. Deforges, and J. Ronsin, "Interleaved S+P Pyramidal Decomposition with Refined Prediction Model," in *ICIP*, October 2005, vol. 2, pp. 750–753.

[5] O. Deforges, M. Babel, and J. Motsch, "The RWHaT+P for an improved lossless multiresolution coding," in *Eusipco*, September 2006.

[6] P. Lux, "A novel set of closed orthogonal functions for picture coding," *Archiv für Electronik und übertragungs Technik*, vol. 31, no. 7, pp. 267–274, 1977.

[7] K. Komatsu and K. Sezaki, "Lossless 2D Discrete Walsh-Hadamard Transform," *Proc. IEEE ICASSP*, May 2001.

[8] G. C. K. Abhayaratne, *Lossless and Nearly Lossless Digital Video Coding*, Ph.D. thesis, University of Bath, 2002.

[9] X. Wu and N. Memon and, "A Context-based, Adaptive, Lossless/Nearly-Lossless Coding Scheme for Continuous-Tone Images (CALIC)," *International Standards Organization working document, ISO/IEC SC29/WG 1/N256*, 1995.