Efficient Pre-Coding Techniques for Wavelet-Based Image Compression*

Detlev Marpe and Hans L. Cycon

Fachhochschule für Technik und Wirtschaft
Berlin, Germany
e-mail: {dmarpe, hcycon}@fhtw-berlin.de

ABSTRACT

The principle of transform coding is a successfully established concept in image compression. In this paper we introduce a coding method using a fast wavelet transform and an uniform quantizer combined with a new framework of pre-coding techniques which are based on the concepts of partitioning, aggregation and conditional coding (PACC). Following these concepts, the data object emerging from the quantizer will be first partitioned into different subsources. Parts of correlations within and between different subsources will then be captured by aggregating homogeneous elements into data structures like run-length codes or zerotrees. By using models based on conditional probabilities we are able to recover correlations between the structures constructed before as well as cross-correlations between different subsources which will be utilized in a final arithmetic coding stage. Experimental results show that our proposed coding methods have a rate-distortion (RD) performance comparable to or even better than the best zerotree-based still image coders in the published literature with the advantage of a less demanding computational complexity. In addition, we propose and evaluate a wavelet-based video coding algorithm which outperforms the very efficient MPEG-4 Video Verification Model (VM 5.1) in both subjective and objective quality.

1 INTRODUCTION

In this paper we introduce a novel framework for the design of effective compression schemes using the discrete wavelet transform (DWT). Wavelet-based image compression schemes combine transform coding with subband coding techniques. Usually it is made up of three modules: first and main part is an invertible transformation which decomposes the signal into a dyadic structured tree of subbands by convolution-decimation operations. It decorrelates mutually dependent parts and performs an energy compaction. The second module is a quantizer which annihilates visually non-relevant information, and the third and final part in this three-stage process consists of an entropy coder which removes statistical redundancies in the stream of symbols generated by the quantizer.

Many (very effective) research activities intend to optimize the transformation and/or the quantization in order to increase the RD-performance of the coder [5, 7, 15]. In this paper, however, we present a coding scheme using a non-sophisticated fast wavelet transform and a simple uniform quantizer with deadzone. We concentrate our optimization activities on the coding part. Actually, we insert a so-called *pre-coder* module which performs a pre-processing of the quantized wavelet coefficients prior to entropy coding. The idea is based on the observation that the wavelet transform performs a localization in phase space (i.e. space-frequency domain) which matches the characteristics of natural images and reveals their scale-invariant features. From a coding point of view, we ask for a mechanism to exploit the redundancy which stems from these characteristics of the wavelet representation. This leads to the proposal of the novel PACC-framework based on the concepts of *partitioning*, *aggregation* and *conditional coding* which provides us with a "toolbox" for constructing of efficient pre-coding techniques.

The organization of the paper is as follows. In an introductory section we review the properties of the DWT and introduce our quantization method. In a subsequent section we state the essential principles of PACC and derive a collection of pre-coding methods from these principles. A more detailed description of two algorithmic realizations will be given as a reference for the remaining part of this paper where we discuss some applications, which demonstrate the excellent performance of our pre-coding methods embedded in a full codec.

*This work was supported by Deutsche Telekom AG, Technologiezentrum Darmstadt, and Deutsche Telekom Berkom GmbH, Berlin, Germany.

2 IMAGE REPRESENTATION AND QUANTIZATION

2.1 Multiresolution Analysis

The framework of multiresolution analysis allows an efficient implementation of the DWT using a perfect reconstruction two-channel filterbank. Given an image I and a filterbank with lowpass-filter H and highpass-filter G, the first step of a wavelet decomposition is performed by repeated application of H and G on rows and columns, subsequently. Due to the separable nature of this filtering process, we get a representation with components in four subbands $\{W_{1,k} | \, k = 0, 1, 2, 3\}$ of different frequency localization. In our notation, the first subscript indicates the (first) *level* of decomposition and the second index denotes the *orientation* of the band according to the four possible combinations of applying H and G to rows and columns (HH, HG, GH, GG). Iterating the row- and column-wise operations of convolution-decimation on the resulting lowest frequency subband $W_{l,0}$ $(l \geq 1)$ yields an unbalanced logarithmic tree-structure of subbands which represents different resolution levels of the input image I.

2.2 Uniform Scalar Quantization

Quantization is the part of our coding scheme where we have to eliminate information in the wavelet transformed image in such a way that under the constraint of a given target bitrate the resulting artifacts in the reconstructed image tend to be below the threshold of visibility. A more elaborate solution of this problem would involve a reliable mathematical model of human visual perception which unfortunately is not yet available. However, starting with the simplest model of an uniform scalar quantizer with an overall stepsize q[1], we keep the option of a perceptual weighting of the "quality factor" q [13]. To absorb those wavelet coefficients, which are essentially related to noise, we have implemented a *deadzone*, i.e. a larger zero bin $[-\tau, \tau]$. The ratio η of zero bin size to stepsize of

$$\eta = \frac{2\tau}{q} = 1.5 \tag{1}$$

was found empirically to be a good choice for all tested image sources at various bitrates. Moreover, following a mean-square error minimizing strategy, the choice of η in (1) has proven to be a good approximation of the optimum, which was characterized analytically under certain statistical assumptions in a recently published work [4].

[1]Note that normalization of the DWT was chosen such that the filter coefficients of H sum up to $\sqrt{2}$.

3 PRE-CODING METHODS

Coding is the task to remove or at least to reduce the statistical redundancy in a given source. Obviously, the efficiency of a coding method relies on an adequate model of the underlying statistics. To design and establish such a model, is what we call pre-coding, and it is reasonable to separate it from the actual (statistical) coding. Assuming some higher degree of statistical dependencies in our given data, a straightforward approach to a model of order N would require to estimate probability distributions (PD) for M^{N+1} different values of a random vector $\mathbf{s} = (s_0, \ldots, s_N)$ with alphabet size M. However, even for a second order model with a small alphabet size $M = 16$, this approach would require to model the amount of 2^{12} PDs, which is practically impossible.

Thus, we have to make some simplifications, while maintaining a maximal degree of statistical dependency. One step in this direction is to divide the source and to reduce the alphabet size with the instrument of partitioning.

3.1 Partitioning

The theoretical basis of partitioning is given by the following two theorems, which were initially introduced in a different context [16] and later discovered to be useful in wavelet-based image coding [11] (proofs are omitted):

(Fixed) Source Partition: *Given a source $\mathcal{S} = \bigcup_i \mathcal{S}_i$, the entropy-rate $\mathcal{R}$ of $\mathcal{S}$ and of its disjoint, nonempty subsources $\mathcal{S}_i$ are related by*

$$\sum_i \mathcal{R}(\mathcal{S}_i) \leq \mathcal{R}(\mathcal{S}), \tag{2}$$

with equality if and only if all $\mathcal{S}_i$ have the same PD.

(Adaptive) Range Partition: *Let the dynamic range of a source $\mathcal{S}$ be given by $\mathcal{A} = \bigcup_i \mathcal{A}_i$, where $\mathcal{A}_i$ are disjoint, nonempty subsets of $\mathcal{A}$ and define the subsources $\tilde{\mathcal{S}}_i = \{s \in \mathcal{S} | \, s = \alpha, \alpha \in \mathcal{A}_i\}$ and the indicator set $\chi = \{x_k | \, x_k = i, \text{ if } s_k \in \mathcal{A}_i\}$, then the total entropy rate is given by*

$$\mathcal{R}(\mathcal{S}) = \sum_i \mathcal{R}(\tilde{\mathcal{S}}_i) + \mathcal{R}(\chi). \tag{3}$$

In our coding scenario, the principle of source partition has one obvious application: we will split the quantized and wavelet transformed image $\hat{c}$ according to its subband structure into subsources $\hat{c}_{l,k}$ with hopefully different PD. Although this first step of partitioning may already decrease the overall (theoretical) rate according to Inequality (2), we will add two iterated steps of an adaptive range partition. This kind of partitioning does not increase the entropy rate (which is the essential interpretation of Eq. (3)), but it allows to disentangle the information. Dividing the range in significant, i.e., non-zero quantized and insignificant, i.e., zero quantized values, will result in a subsource of significant coefficients $\hat{c}_{l,k}^{\mathrm{sig}}$ and a source containing the side-information of the adaptive partition, the so-called *significance map* $\chi_{l,k}$. A further range partition according to the sign of significant coefficients finally yields using Eq. (3)

$$\mathcal{R}(\hat{c}_{l,k}) = \mathcal{R}(\chi_{l,k}) + \mathcal{R}(\hat{c}_{l,k}^{\mathrm{mag}}) + \mathcal{R}(\varsigma_{l,k}), \qquad (4)$$

where the *magnitude map* $\hat{c}_{l,k}^{\mathrm{mag}}$ is the subsource containing the magnitudes of significant coefficients and the *sign map* $\varsigma_{l,k}$ holds the relevant sign information. (See Fig. 3 for an instructive example.)

Eq. (4) tells us that our replacement of $\hat{c}_{l,k}$ with two binary-valued indicator maps $\chi_{l,k}$ and $\varsigma_{l,k}$ and a magnitude map $\hat{c}_{l,k}^{\mathrm{mag}}$ does not alter the lower bound on the attainable coding rate, but, what is more important, it gives us more insight into the way how we can approximate this ultimate bound:

- The interpretation of $\chi_{l,k}$ and $\varsigma_{l,k}$ as bilevel "images" offers us the whole arsenal of lossless bilevel coding methods.
- The interdependence of the different subsources either of different type or of different indices (l, k) should be taken into account.

Analyzing the bilevel maps with respect to this strategy, we observe the occurrence of specific patterns or clusters of states with some kind of coherence across different levels l. In the next section, we will discuss methods to utilize this kind of correlations by aggregating homogeneous elements into quadtrees and alternating sequences of runs.

3.2 Aggregation

3.2.1 Zerotree-Aggregation

Relating insignificant coefficients in the wavelet representation which share the same spatial location but reside in different levels, we can build balanced quadtrees of insignificant coefficients, so-called *zerotrees* (ZT). In the pioneer work of [3, 9], the zerotree data structure has been recognized as an useful tool to exploit the complementary part of self-similar structures in the multiresolution representation. Our approach differs from the previous ones in the way we handle the zerotree root symbol. Given the significance map $\chi_{l,k}$ $(l > 1)$, we examine for each insignificant coefficient (which is not part of a ZT found on a previous level) whether it is a ZT root or not. If it is, we assign a "0"-symbol, else we put in a "1"-symbol (the symbol used for the significant coefficients) and move the so-called *isolated zero* to the magnitude map $\hat{c}_{l,k}^{\mathrm{mag}}$ (cf. Fig. 3).

The result of this zerotree analysis is that we can replace $\chi_{l,k}$ and $\hat{c}_{l,k}^{\mathrm{mag}}$ by a binary-valued *zerotree map* $\chi_{l,k}^{\mathrm{ZT}}$ indicating the positions of ZT roots and of coefficients which are not part of a ZT together with a *modified magnitude map* $\hat{c}_{l,k}^{\mathrm{mm}}$ which includes isolated zeros. This leads to a more compact representation of insignificance at the expense of an enlarged magnitude map.

3.2.2 Run-Length-Aggregation

So far, we have only utilized the correlations between insignificant coefficients across the scales. Analyzing the bilevel zerotree-map band-wise, we observe aggregates of significant coefficients[2] typically representing "edges" or parts of "texture"-elements with a (preferably) definite orientation coinciding with the band's orientation. The easiest way of capturing these aggregates (and its complement) is to use runs of both consecutive "0"- and "1"-symbols along a pre-defined scanning path in accordance with the orientation of a given band. Since the statistics of both types of run-length are usually very different and non-stationary, we propose the use of different adaptive models for each type. The experiments we performed to compare this new hybrid zerotree/run-length approach to the common (bilevel) run-length encoding (RLE) of the significance map yield a considerable coding gain of 5–10% for the proposed method at low to medium bitrates. The main contribution to this gain may be attributed to a better performance especially at high-frequency bands. Skipping the ZT elements, we get there a relatively small number of significantly shorter runs of "0"s with the penalty of a comparatively small overhead in the lower-frequency subbands due to the impact of isolated zeros.

Our analysis of the sign map $\varsigma_{l,k}$ showed that there are locally extended regions of constant sign and characteristic patterns of sign changes. These sign changes are usually due to edges having an orientation with a strong bias towards the orientation k of the given band. This observation motivated the use of a simple one-type RLE for the aggregation of chains of constant sign. Compared to the alternative of leaving the sign map uncoded, we are able to realize with this simple and hence computational less demanding approach a bitrate reduction of a few percent, justifying its incorporation.

[2] Note that we also modify our terminology in the sense that an isolated zero is now called significant.

3.3 Conditional Coding

3.3.1 Template Modeling

As mentioned above, for the purpose of encoding the binary-valued indicator maps we can use all sorts of coding methods developed for the (lossless) compression of bilevel images. Run-length coding as described in the preceding section is one possibility which offers a good trade-off between coding efficiency and complexity. However, sacrificing speed the context-based modeling introduced in [2] and successfully implemented in the JBIG[3] standard offers a more effective coding strategy.

Essentially, this approach is based on a model using conditional probabilities where the conditioning "context" is created with the help of a so-called *template*. A template is usually made up of neighboring elements of the current element to encode. Fig. 1 (a) shows the template, we have designed for the purpose of coding the zerotree map $\chi^{\mathrm{ZT}}_{l,k}$. It is similar to the differential-layer template used in the JBIG standard and covers elements of two levels: 5 surrounding elements of the current one (C) and two neighbors of the "parent" P of C, which allow a "prediction" of the non-causal neighborhood of C. In addition, we adapt the template to the orientation of the band by choosing one of its elements according to the direction of predominant correlations (cf. Fig. 1 (a)).

3.3.2 Conditioning Categories

Further improvements on coding efficiency can be achieved by using conditional probabilities to encode the (modified) magnitude map. Since we established a definite order of processing the subsources band-wise by first coding (and decoding) the zerotree map, we can use this available information for the construction of conditioning categories. Fig. 1 (b) shows the 3×3-window of a current significant coefficient $c = \hat{c}^{\mathrm{mm}}_{l,k}[n, m]$ which is "mapped" on the corresponding part of the zerotree map in order to define the conditioning descriptor $\kappa = \kappa_{l,k}[n, m]$ of c

$$\begin{aligned}\kappa \;=\; & \chi[n-1,m] + \chi[n,m-1] \\ & + \chi[n+1,m] + \chi[n,m+1] \qquad (5) \\ & + ((\chi[n-1,m-1] \vee \chi[n+1,m+1]) \\ & \;\wedge (\chi[n+1,m-1] \vee \chi[n-1,m+1]))\,,\end{aligned}$$

where $\chi = \chi^{\mathrm{ZT}}_{l,k}$ ($\vee, \wedge$: logical "or" resp. "and").

To reduce the overall learning cost in the adaptive arithmetic coder, we restrict the number of conditioning categories to 6 states characterized by the r.h.s of Eq. (5) which we found experimentally to yield best performance.

[3]Joint Bilevel Image Experts Group (ISO committee)

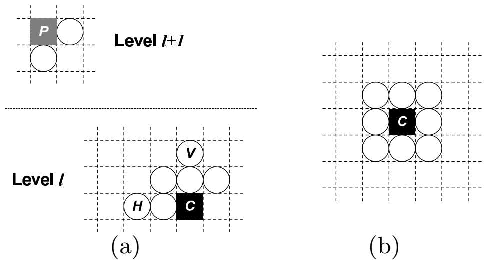

Fig. 1: (a) 7-element orientation-adaptive two-scale template (white circles) for context-based coding of the zerotree map. V and H indicate adaptive elements used for vertical and both horizontal and diagonal bands, respectively. (b) 8-neighborhood used for conditioning of C

4 EXPERIMENTAL RESULTS

Two Algorithms We present a brief description of two algorithmic realizations of our proposed methods. Both variants use a 4-level DWT with either biorthogonal 9/7-tap filter [1] for low-resolution images (such as QCIF video sources) or biorthogonal 10/18-tap filter [10] for high-resolution still images. The way they differ is related to the method of zerotree map coding:

PACC-1: hybrid ZT/RLE of significance map
PACC-2: conditional coding of zerotree map

Except for the lowest frequency band which is coded separately using a DPCM-like procedure, all bands are pre-processed in both coding algorithms using the methods of (range) partition, zerotree aggregation, run-length aggregation of sign map and conditional coding of magnitudes. The resulting symbol streams are piped through an adaptive arithmetic coder [14] which using the appropriately supplied model produces a decodable bitstream[4].

Bitrate	PACC-1		S&P	
(bpp)	PSNR	CPU	PSNR	CPU
0.5	37.20	0.44	37.21	0.74
0.25	34.14	0.31	34.11	0.45
0.2	33.17	0.28	33.15	0.39

Table 1: Evaluation of our algorithm PACC-1 in comparison with S&P [8] for the standard 512×512 Lena image. CPU coding times (sec) are reported for SGI-Indy 150MHz.

Still Image Compression In Table 1 and 2, our results are compared against those obtained with two of the best published zerotree-based still image coders, the SFQ-coder by Xiong et al. [15] and the rate-scalable coder of Said and Pearlman (S&P) [8]. Our PACC-1 coder shows RD-performance similar to the S&P-coder with a significantly, i.e, 20–45% lower complexity (cf. Table 1).

[4]Side information has been taken into account for rate calculations.

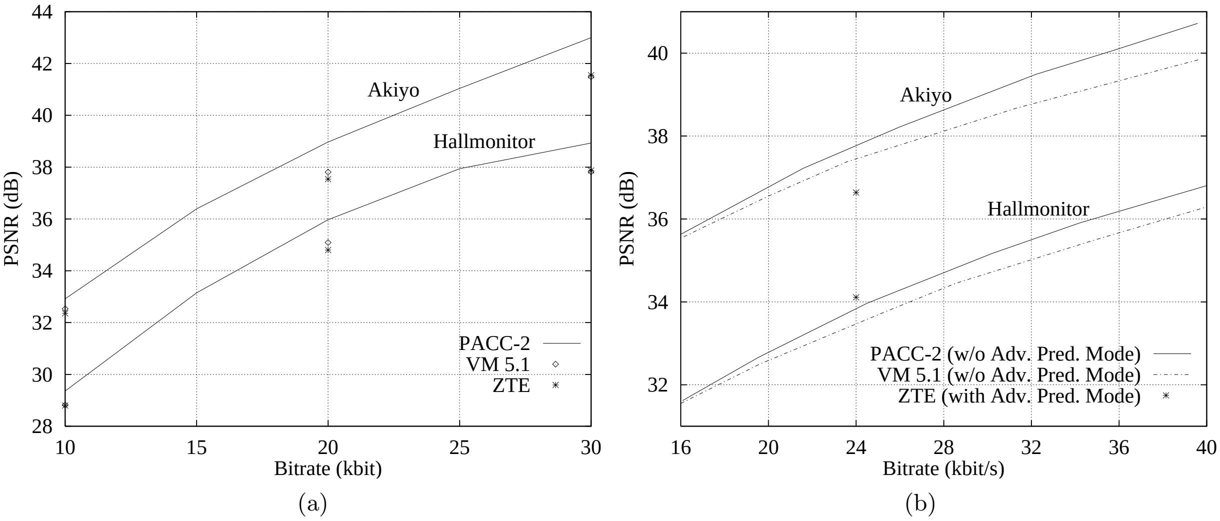

Fig. 2: Bitrate vs. Y-PSNR:(a) I-frame and (b) I- plus P-frame coding of Akiyo and Hallmonitor (10 frames/sec)

Better results can be obtained with the PACC-2 coder at the expense of an approximately doubled coding time compared to PACC-1. This computational requirement is quite demanding for additional savings in bitrate of 5% ($\approx$ 0.2 dB PSNR), but it is presumably far less than the (not reported) execution time of the equally performing SFQ-coder (see Table 2) which is based on complex on-line computations to determine RD-optimal quantization parameters.

Bitrate (bpp)	*Lena*		*Goldhill*	
	PACC-2	SFQ	PACC-2	SFQ
0.5	37.37	37.36	33.32	33.37
0.25	34.36	34.33	30.68	30.71
0.2	33.41	33.32	29.96	29.86

Table 2: RD-Performance comparison of our algorithm PACC-2 and SFQ-coder [15]. Coding results are reported in PSNR (dB) for the 512×512 Lena and Goldhill images.

Video Coding For the purpose of video coding, it is crucial to take advantage of the correlation in the temporal dimension. Therefore, we will supply our coding algorithms with a temporal predictive feedback loop, where prediction is performed using a block motion estimation similar to that of H.263[5]. However, it is well-known that a box-function shaped window function has a transfer function with slow decay and hence introduces annoying high-frequency components in the resulting prediction residual (P-frame). To overcome this drawback, we use an overlapping and smoothing window investigated in [12]. In addition, a simple but efficient technique allows further utilization of temporal redundancy in the way adaption of probability models will be performed, i.e., consecutive P-frames as well as consecutive motion vector fields will be encoded using the updated models of the related previous P-frame and motion field, respectively. Furthermore, we modify the zerotree analysis for P-frame coding to include the lowest frequency band.

For our experiments MPEG-4 test sequences at QCIF resolution with 4:1:1 YUV color representation and 30 frames/sec have been temporally subsampled to yield a frame rate of 10 frames/sec. Results for coding of initial I-frame alone and averaged results for coding of the entire test sequence (1 I- and 99 following P-frames) are shown in the graphs of Fig. 2 (a) and (b), respectively. We compare our results with those obtained with a greatly enhanced H.263 coder used for verification and evaluation of MPEG-4 tools and algorithms, so-called *VM*[6] (Version 5.1), and with a zerotree-based video coder (ZTE) [6].

Fig. 2 (a) shows that I-frame coding of PACC-2 outperforms that of both VM and ZTE by 0.5–1.5 dB PSNR. The evaluation of P-frame coding, however, is somehow involved since different temporal prediction schemes were used. The (only) two available results for ZTE plotted in Fig. 2 (b) were obtained with an improved technique similar to the *Advanced Prediction Mode* of H.263 which uses four additional 8×8 subblocks for a better prediction of moving small-scale structures. This is obviously critical for "Hallmonitor" showing two walking figures captured from a distant monitoring camera. Nevertheless, with primal focus on coding efficiency, the results plotted in Fig. 2 (b) demonstrate a consistently superior RD-performance of PACC-2 compared with VM.

[5]ITU-T Draft Recommendation "Video Coding for Low Bitrate Communication", Dec. 1995.

[6]MPEG-4 Video Verification Model

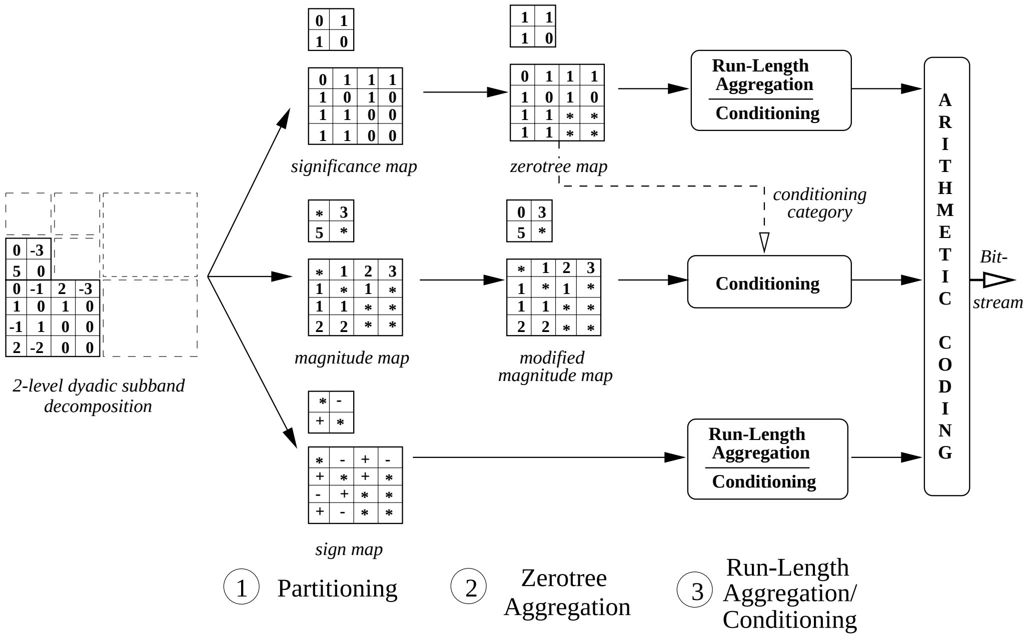

Fig. 3: An example to illustrate the 3-stage pre-coding process. Incoming object is a 2-level quantized wavelet decomposition of a 8×8 "toy image". Note that the schematic representation displays only processing of the "horizontal" orientation ($k = 1$). Partitioning (Step 1) yields 2 bands per orientation ($k > 0$) and 3 parts for each band: a significance map, indicating the position of significant coefficients, a magnitude map and a sign map with entries for each non-zero coefficient (zero coeff. = "*"). Zerotree analysis (Step 2) is only performed for the upper level ($l > 1$), which results in one ZT root depicted by a "0"-symbol at the lower right zero coefficient (descendants on the lower level are marked by a "*" in the zerotree map). Note that the isolated zero (upper left) moves to the magnitude map and hence labeled significant ("1"-symbol) in the zerotree map. For the details of Step 3, the reader is referred to Sec. 3.2.2 and 3.3.

5 CONCLUSION

In this paper we presented the novel and unifying coding strategy of partitioning, aggregation and conditional coding. We demonstrated how this framework can be used for the design of efficient and fast coding schemes for still and moving pictures. Our coding results which are among the best currently available confirmed our belief that this new framework with its not yet fully exploited potential is well suited to the wavelet paradigm of joint space-frequency localization.

6 REFERENCES

[1] A. Cohen, I. Daubechies and J.-C. Feauveau, "Biorthogonal Bases of Compactly Supported Wavelets", *Comm. on Pure and Appl. Math.*, Vol. 45, pp. 485–560, 1992.

[2] G. G. Langdon and J. J. Rissanen, "Compression of Black-White Images with Arithmetic Coding", *IEEE Trans. on Comm.*, Vol. 29, No.6, pp.858–867, 1981.

[3] A. Lewis and G. Knowles, "Image Compression Using the 2D Wavelet Transform", *IEEE Trans. on Image Proc. (IP)*, Vol. 1, No. 2, pp. 244–250, April 1992.

[4] S. Mallat and F. Falzon, "Understanding Image Transform Codes", *Proc. SPIE Aerospace Conf.*, Orlando, April 1997.

[5] D. Marpe, H. L. Cycon and W. Li, "A Complexity Constrained Best-Basis Wavelet Packet Algorithm for Image Compression", to appear in *Applied and Computational Harmonic Analysis*, 1997.

[6] S. A. Martucci, I. Sodagar, T. Chiang and Y. Zhang, "A Zerotree Wavelet Video Coder", *IEEE Trans. on Circuits and Systems for Video Technology (CSVT)*, Vol. 7, No. 1, pp. 109–118, Feb. 1997.

[7] K. Ramchandran and M. Vetterli, "Best Wavelet Packet Bases in a Rate-distortion Sense", *IEEE Trans. on Image Proc.*, Vol. 2, No. 2, pp. 160–175, April 1993.

[8] A. Said and W. A. Pearlman, "A New Fast and Efficient Image Codec Based on Set Partitioning in Hierarchical Trees", *IEEE Trans. on CSVT*, Vol. 6, No. 3, pp.243–250, 1996.

[9] J. M. Shapiro, "Embedded Image Coding Using Zerotrees of Wavelet Coefficients", *IEEE Trans. on Signal Proc.*, Vol. 41, No. 12, pp. 3445–3462, Dec. 1993.

[10] M. J. Tsai, J. Villasenor and F. Chen, "Stack-run Image Coding", *IEEE Trans. on CSVT*, Vol. 6, pp. 519–521, Oct. 1996.

[11] Y. Wang and E. Salari,"The Post Wavelet Transform Redundancy and its Reduction Techniques for Image Compression", *Proc. SPIE*, Vol. 2418, pp.164–173, 1995.

[12] H. Watanabe and S. Singhal, "Windowed motion compensation", *Proc. SPIE* Vol. 1605, pp.582–589, Nov. 1991.

[13] A. B. Watson, G. Y. Yang, J. A. Solomon and J. Villasenor, "Visibility of Wavelet Quantization Noise", to appear in *IEEE Trans. on IP*, 1997.

[14] I. Witten, R. Neal and J. Cleary, "Arithmetic Coding for Data Compression", *Comm. ACM*, Vol. 30, pp. 520–540, June 1987.

[15] Z. Xiong, K. Ramchandran and M. T. Orchard, "Space-frequency Quantization for Wavelet Image Coding", to appear in *IEEE Trans. on IP*, 1997.

[16] Y. Huang, H. M. Dreizen and N. P. Galatsanos, "Prioritized DCT for Compression and Progressive Transmission of Images", *IEEE Trans. on IP*, Vol. 2, No. 4, pp. 477–487, Oct. 1992.