MODIFICATIONS OF UNIFORM QUANTIZATION APPLIED IN WAVELET CODER

Artur Przelaskowski

Institute of Radioelectronics, Warsaw University of Technology
ul. Nowowiejska 15/19, 00-665 Warszawa, Poland

Abstract

An algorithm of wavelet domain data quantization aimed at improving compression efficiency is presented. Threshold data selection is proposed as more effective uniform quantization modification than zero-zone increasing. To fit adaptively threshold value to local image features, the estimation of significance expectation for each wavelet coefficient was included into thresholding procedure. The remaining data are uniformly quantized without any changes of bin boundaries. As a result, more effective low-cost quantization scheme was constructed. It allows significantly increase compression efficiency of images. Experimental Rate-Distortion curve shows the same distortion for decreased bit rates even up to 20% in comparison to standard uniform quantization. Such quantization technique was applied in wavelet coder with optimised schemes of decomposition and zerotree based coding. Its compression efficiency is competitive with the most efficient methods across all natural images tested.

1. INTRODUCTION

A quantization procedure is fundamental in lossy compression, also for wavelet-based techniques. Space-scale data coefficient localisation in wavelet domain reflects nonstationary nature of images and determines profitable conditions for nontrivial solutions of quantization problem. Global, stationary models, assumptions of i.i.d (independent and identically distributed) signals and others modelling tools are often too coarse to capture the character of image plane features, containing important information which should be preserved in compression.

Considerations on entropy constrained scalar quantization prove that uniform quantizers are optimal for high bit rate compression [1]. But established high resolution quantization hypothesis is not valid in case of low bit rate compression (bit rates less than 1.0 bpp). Quantization bins are too large and probability density of quantized random source can not be considered as approximately constant. However, Farvardin and Modestino [2] proved that even though the high resolution assumption may not hold, for a large class of probability distribution including Generalised Gaussian Distribution (GGD), the uniform quantizer yields a distortion rate that is close to the optimal quantizer if the number of quantization bins is large enough. In this case uniform threshold quantization (UTQ) has been proved to perform very close to the optimal entropy constrained quantizers for a wide class of memoryless sources. The quantizers, which have the infinite number of levels and equal step width belong to UTQ class and modifications of UTQ are used in the most efficient compression algorithms [3][4][5].

The purpose of our research is to increase efficiency of quantization scheme by improving uniform quantization procedure and exploiting and space-scale data dependencies. The optimisation of quantization process is realised with respect to adaptive threshold data selection across scale and space.

2. STATISTICAL MODELLING AND QUANTIZATION OF WAVELET COEFFICIENTS

Generally, the statistical prior model, even if it captures the variations in data dependencies and appearance only partially, can be substantially beneficial for image compression because of fitting proper quantization scheme and consequently encoding procedure in compression algorithm. But it is very difficult to infer even complex probability density function modelling data dependencies because of high dimensionality of digital images. Thus it is essential to simplify the problem by reducing the dimensionality of compressed data space by linear, often unitary transformation of original image data. In other words, converting global data inter-dependencies into small local data dependence could be very useful in statistical description of compressed data.

Marginal data distributions are often used to characterise the statistical properties of wavelet coefficients. These models are based on the assumption that the data within a subband are i.i.d. It was noticed, empirically and theoretically, that the marginal statistics $p(x)$ of wavelet domain data are highly non-Gaussian. Zero-mean Generalized Laplacian distribution was found to be a good description of marginal densities [6][7]. LoPresto [3] proposed zero-mean GGD density function. Each subband is viewed as a zero mean GGD source with a standard deviation that is slowly varying function of the spatial coefficient localisation. The flexibility of GGD shape modelling allows for the efficient capture of the diverse statistics of different wavelet bands for various images and target bit rates caused by fixed slope λ of Rate-Distortion (R-D) characteristics. Each coefficient variance is estimated by means of Maximum Likelihood Estimate 'on the fly' based on local context. The assumption that each coefficient of causal context (defined by local causal windows 3×3 and 5×5) is drawn independently from the GGD is used. Backward and forward adaptation scheme is used in final estimation of the model of pixel statistics derived from GGD parameterised source. Novel Mallat and Falzon considerations [1] suggest rational instead of exponential decay of $p(x)$:

$$p(x) \propto x^{-1-\frac{1}{\nu}} \tag{1}$$

for x large enough and exponential decay well-modelled by GGD when x is small.

The weakness of such models is generally related to the assumption of data independence in wavelet domain. The coefficients of wavelet decomposition are found to be fairly well decorrelated but they are not independent. Some data dependence in spatial subband domain (e.g. in edge localisation) and inter-band parent-children relations is noticed. The joint histograms presented by Simoncelli [7] show that high-magnitude parent-children coefficients are dependent. The conditional expectation $\mathrm{E}(C \mid P)$ is approximately proportional to P: $\mathrm{E}(C \mid P) \propto P$, where parent random variable P and child random variable C are defined by sets of coefficient magnitudes. Furthermore, the qualitative character of these statistical relationships for magnitudes of coefficients at adjacent spatial localisation was found. Thus joint magnitude statistics to model local spatial and inter-scale data dependence is used in our modification of dead-zone UTQ.

We examine linear prediction for coefficient magnitude estimation (*ME*): $ME_i^K \doteq \sum_{k=1}^{K} \alpha_k m_{i,k}$, where the adjacent (in space - 3×3 causal window and parent node) coefficient magnitude set $\{m_{i,k}\}$ is used. The magnitude value $m_i = |c_i|$ is conditioned upon the *ME* for each significant coefficient c_i. The weights α_k are chosen to minimise a mean square error of real magnitude values approximation.

The 1st order conditional model $P(m_i \mid ME_i^K)$ is used instead of conditional probability model of *K*-order $P(m_i \mid m_{i,1},...,m_{i,K})$ because of context dilution. Such context quantization can be continued by quantization of *ME's*. The number of levels was decreased about 30% to improve statistical model. Therefore the conditional probability model $P(m_i \mid Q(ME_i^K))$ can increase conditional probability of alphabet symbols in comparison to simpler first order model $P(m_i \mid m_{i-1})$. It captures the majority of mutual information between such linear predictor and coefficient magnitude values.

The optimal entropy constrained scalar quantizer for low bit rates is nearly uniform but not uniform. Normalised histograms of wavelet coefficients for different subbands and images show that probability density estimate *p*(*x*) has important variations and faster decay for small *x* close to zero. Hence non-zero quantization bins have the same size Δ but the zero bin must be modified. Its width $[-\tau, \tau]$ is larger than $[-\Delta/2, \Delta/2]$. The zero bin ratio $\eta = \frac{\tau}{\Delta}$ is a parameter that must be adjusted to optimise compression algorithm in R-D sense. Larger zero bin, called dead-zone, reduces these wavelet coefficients which are essentially related to noise. Both experimental [4] and theoretical [1] ways of optimal η value estimation were successfully applied. Such dead-zone UTQ (DUTQ) scheme is considered as the most optimal for wavelet compression applications. We suggest the modification of such scheme by incorporating adaptive dead-zone modulation based on joint statistical models of local data dependencies.

LoPresto optimised η by fitting marginal distribution models to data features and selecting an optimal R-D curve. Mallat and Falzon presented the theoretical zero bin ratio evaluation, where the following relation can be stated:

$$\eta \propto \sqrt{\frac{R_1}{M}}; \qquad (2)$$

R_1 - total number of bits of the encoded quantized significant coefficients, *M* - number of significant coefficients. We approximated relation (2) for local optimisation of η_i. The R_1 value depends on many factors of encoding algorithm and modelling of all type dependencies is too complex, but one important note can be stated: R_1 is proportional to the conditional entropy of data source characterised by conditional probability model

$P(m_i \,|\, Q(ME_i^K))$. Because similar probability model is used in arithmetic encoding of quantized coefficients, lower conditional probability of symbols of source alphabet means poorer coding efficiency and increased bit rate of final code. Increased value of h_i should reduce unsuitability of conditional model to improve compression effectiveness.

Data significance s_i of coefficient c_i in relation to threshold T (or t as the width of zero bin) means that $s_i = 1 \Leftrightarrow c_i \geq T$ and oppositely, insignificance means $s_i = 0 \Leftrightarrow c_i < T$. We used the status of surrounding data $\{s_{i,l}\}$ in causal context (order L) of current and higher level to estimate significance expectation for coefficient c_i :

$$\mathrm{E}(s_i) \doteq \tfrac{1}{L} \sum_{l=1}^{L} s_{i,l} \ . \tag{3}$$

Locally, we can state that $M \propto \mathrm{E}(s_i)$. Profits of threshold increase are mainly the result of the removal of encoding inefficiency of significant coefficients surrounded by insignificant pixels. Information content expressed in coefficient value does not make up for the costs of such value encoding (small slope of R-D curve). Significance expectation can be considered as local area activity estimation to improve R-D performance of the coder. Finally, we state the following expression for adaptive zero bin ratio modification:

$$h_i \propto \sqrt{\frac{-\sum_l P(m_l \,|\, Q(ME_i^K)) \log_2 P(m_l \,|\, Q(ME_i^K))}{\mathrm{E}(s_i)}} \ , \tag{4}$$

where l is an index of symbol of current magnitude source alphabet. Such quantization procedure is called adaptive DUTQ (ADUTQ). Let us consider the problem of significance estimation in greater detail in the next section.

3. THRESHOLD DATA SELECTION IN QUANTIZATION ALGORITHM

The most efficient coders are based on zerotree structures, which describe and exploit the relationships between the wavelet coefficients across subbands. Simple and useful model of zerotree structures could be disturbed by hypothetical situation, where only a small group of significant pixels (even one) as an edge sign appear in the middle of large low activity area which could be efficiently covered by well fitted zerotree structure. Let us assume that a single pixel is located at the bottom of great zerotree extended across several tree levels. It is possible to happen in first steps of iterative EZW-like algorithm (e.g. [11]). The profits of very efficient coding of a large set of data by single index of large zerotree are lost in that case. A few smaller zerotrees and single coefficients are probably coded instead, with additional code bits. In rate-distortion (R-D) sense to ignore this single pixel is more efficient than to preserve this pixel and lose coding efficiency. Entire wavelet coefficient selection is a good tool to eliminate these undesirable single significant pixels in large low activity areas.

Thus to determine the circumstances under which significant coefficients should be ignored is the main problem of compression efficiency improvement. This issue could be considered more formal in R-D sense but the problem of finding the space-scale balance parameters for optimal compression efficiency at each case is not solved yet. Iterative or interactive procedure of space-frequency quantization must be applied to optimise compression scheme in R-D sense because of a balance between ignoring and preserving pixels considered as optimal zerotree pruning. Our problem can be stated simply as:

$$\min_{\{\Delta \in Q; \tau \subseteq \hat{\mathbf{O}}\}} D(\Delta, \tau) \text{ subject to } BR(\Delta, \tau) = BR_d , \tag{5}$$

where $D(\Delta, \tau)$ is a distortion associated with quantizer choice (Δ, τ), Q represents a finite set of all admissible scalar quantizer choices, $\mathbf{T}$ is a full depth decomposition tree and τ is pruned subtree. The shape of pruned subtrees is a function of threshold T in adaptive data selection $\tau(T)$.

Better significance characteristic is possible in completed UTQ scheme. We modified typical dead-zone UTQ scheme. Increased dead-zone was replaced with threshold data selection without any modification of scalar uniform quantization. Thus this algorithm can be considered as entire coefficient selection with threshold value exceeding $\Delta/2$ and quantization of remaining coefficients with constant bin size Δ (threshold data selection and uniform quantization - TSUQ). It is explained in figure 1. The benefit of applying such method is clear: data selection is not repeated in decoder so it could be done without obeying causality. Full size context can be applied in estimation of threshold value.

Additionally, considerations on proper η_i selection are extended on the conditions of threshold T_i optimisation for each coefficient. Small level of T_i increase (less than 0.5Δ) is assumed. According to (5), it is clear that threshold value for each coefficient is a function of step size Δ and currently pruned subtree τ_i: $T_i = f(\Delta, \tau^{\Delta}_{T_0, T_1, \ldots, T_{i-1}},) = f(\Delta, \tau_i)$. Causal part of this tree is provided by thresholding of previous coefficients with $T_0, T_1, \ldots, T_{i-1}$, and noncausal part of tree is defined by Δ. Moreover, we construct an estimator of significance expectation similar to (3), but it is based of L-order context shape specified by decomposition tree $\mathbf{T}$ (parent-children relations) as a function g of τ_i:

$$\bar{s}_i = E[s(\tau_i)] = g(s_i^{\mathbf{T}}) = \sum_{l=1}^{L} s_{i,l}^{\mathbf{T}} . \tag{6}$$

This estimator reflects the shape of subtree τ_i as encoding efficiency indicator in considered R-D optimisation. Hence, the value of T_i should be redefined as a function of data significance $T_i = f(\Delta, \bar{s}_i)$. But the question is about function $f(\cdot)$. Our approximation of optimal $f(\cdot)$ for practical algorithm is as follows. Taking the equation (4) we can replace conditional entropy expression with significance estimator (6) because of: a) zerotree -based structure of lossless encoder, b) coding of magnitudes as successive bit planes, which could be treated as significance maps, c) unchanged bin boundaries and reconstruction levels, and decreased entropy only by significance-to-insignificance coefficient transients. Because thresholding mostly influences significance map, we can approach threshold function by stronger dependence on $\bar{s}_i$: $T_i \propto \Delta(1-\bar{s}_i)^2$. Finally, we concluded the following expression on adaptive threshold modification, verified experimentally:

$$T_i \cong \frac{\Delta}{2}\left[1 + t(1-\bar{s}_i)^2\right], \qquad (7)$$

where t - constant fitted as forward adaptation of image or successive subbands characteristics.

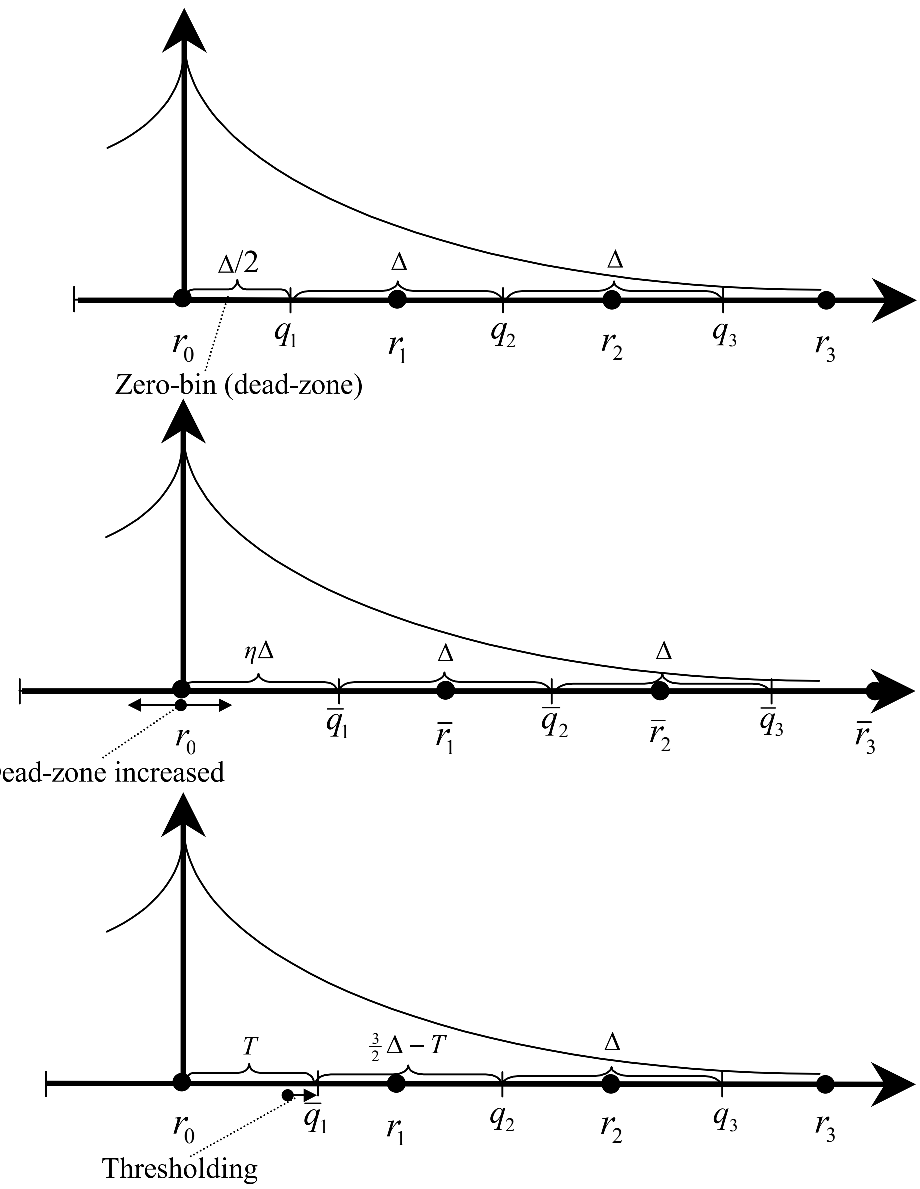

Figure 1. The positive half of the UTQ with step size Δ (top), DUTQ with zero bin ratio η (middle) and TSUQ with threshold T (bottom). q_i's and r_i's are bin boundaries and reconstruction levels, respectively. $\bar{q}_i$'s and $\bar{r}_i$ are changed points in modifications of UTQ.

Based on significance estimator and TSUQ scheme, quantization procedure called adaptive TSUQ (ATSUQ) for each wavelet coefficient c_i is as follows (all computations are low-cost):

- Calculate the shape of current pruned subtree τ_i according to step size Δ and set of prior threshold values $T_0, T_1, ..., T_{i-1}$.
- Evaluate the significance expectation based on adjacent pixels in decomposition tree (other children and parent) according to (6).
- Compute final threshold value T_i defined by formula (7) and select correct coefficient as significant with relation to T_i or insignificant.
- Perform uniform quantization with step size Δ in case of coefficient verification as significant.

4. WAVELET CODER

To implement and test described ATSUQ procedure, we constructed wavelet coder with some optimization of wavelet decomposition and quantized coefficients coding scheme. We also tested different version of such coder to consider compatibility with JPEG2000 paradigm.

4.1 Filters and decomposition

We applied efficient filter banks, known from literature. More compact energy distribution in space-scale domain does not produce separated single points in large smooth area. Thus data covering zerotree structure system could be made more efficient. To optimise wavelet image decomposition in R-D sense one must chose a wavelet basis that gives precise approximation of an image with few wavelet coefficients compactly located in space across different scales (not spread in spatial domain around edges). As a result, a small approximation error and efficient significance map description by the zerotrees is maintained. Additionally, we tested different decomposition schemes: dyadic, standard, uniform, uniform plus dyadic etc. to improve compression efficiency.

4.2 Coding scheme

It is based on zerotree structure built from the top to the bottom of hierarchical wavelet tree. Some simple statistical modelling for arithmetic coder was applied. Magnitudes, signs and significance map were coded separately. We built progressive coding schemes in quality- and in resolution order. The embedded version of coder with ATSUQ and set partitioning similar to SPIHT was constructed.

4.3 Progressive and embedding realisation

Two problems are the most difficult to solve in embedded coding scheme: iterative optimisation of threshold data selection for fixed target bit rate and R-D optimised code generation for all successive lower bit rates until fixed bit rate is finally approached. Threshold data selection and data quantization is scaled as a function of target bit rate.

Progressive in quality encoding is realised as successive data selection for thresholds built on the base of entire step size Δ, next Δ/2, Δ/4 etc. Entire step size in concluded as the extension of Wintz and Kurtenbach [8] considerations:

$$\Delta = k \cdot \left\{ BR_d + \frac{2}{\ln 10} \left\{ \ln \sigma^2 - \frac{1}{N} \sum_{n=1}^{N} w_n \ln \sigma_n^2 \right\} \right\}, \tag{8}$$

where: N - number of subbands, σ_n^2 - coefficient variance estimation for subband n, σ^2 - all coefficients' variance estimation, w_n - weight of subband n closely related to scale, BR_d - average desired bit rate per sample. Factor k could be established as constant or interactively fitted to optimal compression in R-D sense.

Wavelet coder progressive in resolution uses different data ordering. All quantized data are coded from the top to the bottom of decomposition tree by following decomposition levels as finer resolution version of compressed image instead of successive bit planes encoding across a whole image. Output data stream is an embedded code in resolution sense. The iterative procedure is needed to meet exactly the target bit rate constraint. Switching between quality and resolution oriented decoding is possible. Data stream encoded in resolution mode could be decoded in successive bit plan order to achieve quality growing reconstruction of images.

5. RESULTS AND CONCLUSIONS

The efficiency evaluation of considered data models and quantization scheme applied to wavelet-based image compression was the subject of the conducted tests. Natural test images: Lenna, Barbara and Goldhill ($512 \times 512 \times 8$bit) and PSNR as quality measure were used. Figure 2 and table 1 show the effectiveness of quantization scheme modification. The compression efficiency evaluation of wavelet coder is reported in table 2.

Table 1. Comparison of different wavelet coefficients' quantization procedures in wavelet-based compression scheme. PSNR values for 0.25 and 0.5 bpp (bits per pixel) are presented. Filter banks from [9] and 6-levels dyadic decomposition were used.

Quantization Procedure	Lenna		Barbara		Goldhill	
	0.25	0.5	0.25	0.5	0.25	0.5
UTQ	34.03	37.00	27.74	31.77	30.20	32.85
DUTQ	34.32	37.35	28.10	32.08	30.60	33.24
TSUQ	34.39	37.40	28.16	32.15	30.64	33.31
ADUTQ	34.47	37.48	28.24	32.23	30.74	33.38
ATSUQ	34.55	37.57	28.40	32.37	30.78	33.46

The benefits of quantization modifications are clear. Experimental R-D curve shows the same distortion for decreased bit rate up to 20% for presented quantization method in comparison to UTQ. Moreover, ATSUQ gave up to 0.3 dB of PSNR improvement in comparison to TSUQ, up to 0.15dB in comparison to ADUTQ, and up to 0.7 dB in comparison to UTQ. TSUQ was more effective than DUSQ for all images and bit rates. Mean improvement is 0.06 dB of PSNR.

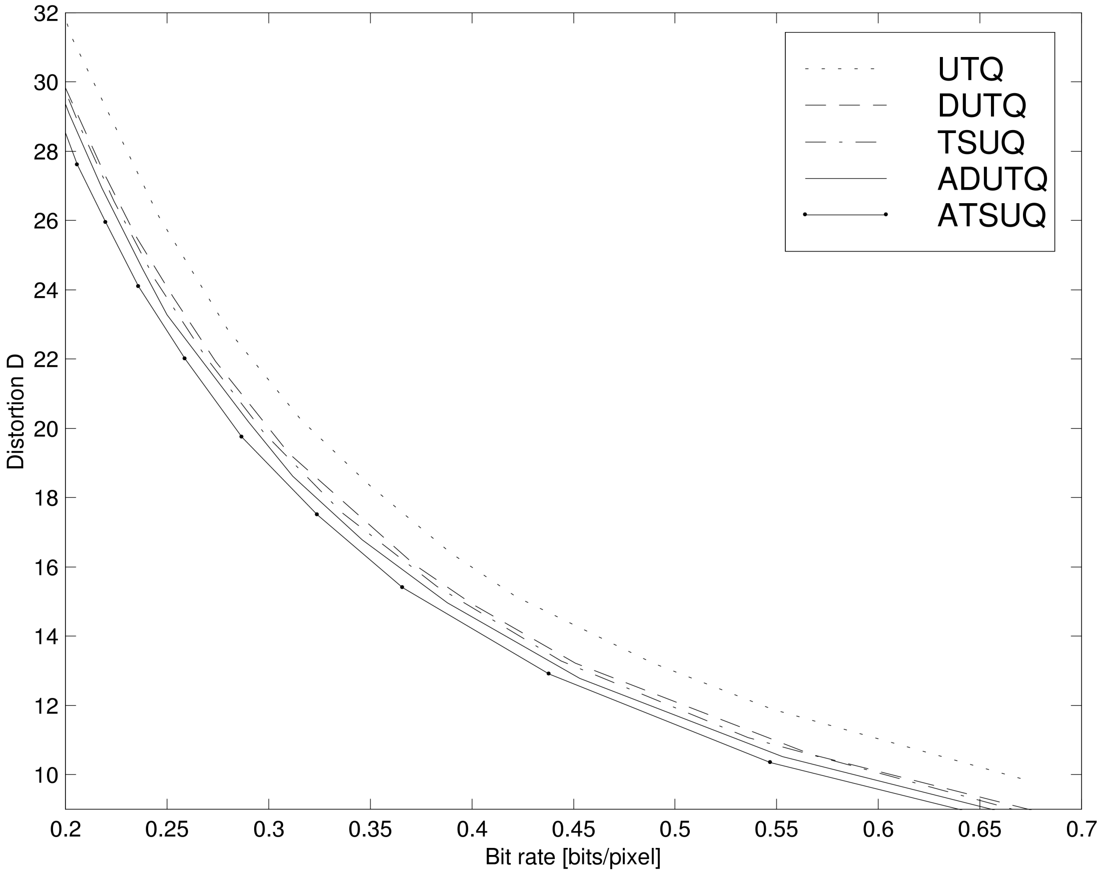

Figure 2. Empirical rate-distortion curve of UTQ, DUTQ, TSUQ, ADUTQ and ATSUQ applied in wavelet coder. Lena was coded and MSE, as distortion D measure was used.

Table 2. The evaluation of compression efficiency of presented wavelet coder. Optimal filter banks ([9] - for Lena and Goldhill, [10] - for Barbara) and decomposition scheme (standard for Lena and Goldhill, 3 levels of uniform plus dyadic for Barbara) were used. SPIHT[11], SFQ[12], C/B[10], PACC[4], PC-AUTQ[5] and EQ[3] are the reference.

Technique	Lena		Barbara		Goldhill	
	0.25	0.5	0.25	0.5	0.25	0.5
SPIHT	34.11	37.21	29.36	33.07	30.56	33.13
SFQ	34.35	37.41	29.67	33.51	30.71	33.37
C/B	34.57	37.52	28.75	32.64	30.80	33.53
PACC	34.53	37.51	28.65	32.54	30.84	33.51
PC-AUTQ	34.46	37.56	-	-	30.78	33.46
EQ	34.57	37.68	-	-	30.76	33.42
Wavelet coder with ATSUQ	34.60	37.58	30.17	33.95	30.99	33.60

Final compression efficiency of wavelet coder with ATSUQ is significantly better than other coders for Barbara (improvement is up to 0.5 dB), slightly better for Goldhill (up to 0.15 dB) and comparable for Lena. Two versions of wavelet coder (progressive in resolution and quality) have comparable compression efficiency results (maximum difference is 0.05 dB of PSNR).

The removal of 'unusual' information in lossy manner can be optimised by proper quantization algorithm design in compression scheme. Presented quantization scheme allows achieving high effectiveness of wavelet compression algorithm.

Scale-space data characteristic after wavelet decomposition is utilised in further data processing and redundancy removal. Entire threshold data selection followed by UTQ was proposed instead of UTQ with increased dead-zone. As a result, two advantages were proved: possibility of noncausal modelling of wavelet coefficients and increased quantization efficiency. Thus, the improvement of quantization scheme by applying data dependency models in wavelet domain was notified. Significance estimation on the base of zerotree structure to combine quantization and coding optimisation is applied.

Adaptive threshold modification in forward and backward manner increased final efficiency of a whole lossy compression algorithm by low computational costs operations. In comparison to more complex wavelet coders with great computational cost of performance, the coder described in this paper is competitive in terms of achieved bit rates across tested images. This coder outperforms reported results in most cases and can be realised in different resolution or quality oriented way.

6. REFERENCES

[1] S. Mallat and F. Falzon, "Analysis of Low Bit Rate Image Transform Coding," *IEEE Tran. Sigal Proc.,* April 1998.

[2] N. Farvardin and J.W. Modestino, "Optimum quantizer performance for a class of non-gaussian memoryless sources," *IEEE Trans. Inform. Theory*, 30:485-497, 1984.

[3] S.M. LoPresto, K. Ramchandran, and M.T. Orchard, "Image Coding based on Mixture Modeling of Wavelet Coefficients and a Fast Estimation-Quantization Framework," *IEEE Data Compression Conference '97 Proc*, 1997, pages 221-230.

[4] D. Marpe, and H.L. Cycon, "Efficient Pre-Coding Techniques for Wavelet-Based Image Compression," *PCS,* Berlin, 1997.

[5] Y. Yoo, A. Ortega, and B. Yu, "Progressive Classification and Adaptive Quantization of Image Subbands," submitted to *IEEE Tran. Image Processing*, 1997.

[6] S.G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Pat. Anal. Mach. Intell.*, 11:674-693, 1989.

[7] E.P. Simoncelli, "Modeling the Statistics of Images in the Wavelet Domain", *Proc. SPIE, 44th Annual Meeting*, vol. 3813, 1999.

[8] P.A. Wintz, and A.J. Kurtenbach, "Waveform error control in PCM telemetry," IEEE *Trans. Inform. Theory*, IT-14:650-661, Sept. 1968.

[9] D. Wei, H-T. Pai, and A.C. Bovik, "Antisymmetric biorthogonal coiflets for Image coding," *Proc of IEEE Intern. Conference on Image Process.,* 2:282-286, 1998.

[10] C. Chrysafis, and A. Ortega, "Efficient context-based entropy coding for lossy wavelet image compression," DCC, Snowbird, UT, 1997.

[11] A. Said, and W.A. Pearlman, "A new fast and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circ. & Syst. Video. Tech.*, 6:243-250, 1996.

[12] Z. Xiong, K. Ramchandran, and M.T. Orchard, "Space-frequency quantization for wavelet image coding," *IEEE Trans. Image Process.*, 6:677-693, 1997.