On the Optimality of Embedded Deadzone Scalar-Quantizers for Wavelet-based L-infinite-constrained Image Coding

Alin Alecu,1) Adrian Munteanu,1) Jan Cornelis,1) Steven Dewitte1,3), Peter Schelkens,1,2)

1 Electronics and Information Processing Department, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium. Email: {aalucu, acmuntea, pschelke, jpcornel, stdewitt}@etro.vub.ac.be}, Phone: +32-2-629-3951, Fax: +32-2-629-2883.
2 Postdoctoral fellowship with Fund for Scientific Research - Flanders (FWO), Belgium.
3 Royal Meteorological Institute of Belgium (RMIB), Ringlaan 3, 1180 Brussels, Belgium

Abstract: In wavelet-based $L_\infty$-constrained embedded coding, the embedded bit-stream is truncated at the bit-rate that corresponds to a guaranteed, user-defined distortion bound. The paper analyses the optimality of embedded deadzone scalar-quantizers for high-rate $L_\infty$-constrained scalable wavelet-based coding of images. A rate-distortion model applicable to the family of embedded deadzone scalar-quantizers is derived; the model is validated experimentally and conclusions are drawn regarding the optimal subband-quantizer instantiations. The optimal quantizers are employed in a coding algorithm that retains the capabilities of wavelet-based codecs while allowing for a fully embedded $L_\infty$-oriented bit-stream.

I. INTRODUCTION

Several applications require compression schemes that guarantee constraints on the local reconstruction (pixel) error rather than on the average error. The local reconstruction error is referred to as the MAXimum Absolute Difference (MAXAD) between the pixel values in the original and reconstructed images. Various methods for $L_\infty$-distortion constrained compression have been proposed in literature, some operating in the transform domain [1], others in the image domain [2], while others propose a hybrid bit-stream between the two [3]. We have recently proposed in [4, 5] a wavelet-based $L_\infty$-constrained scalable image-coding technique that generates
a fully embedded $L_\infty$-oriented bit-stream, while retaining the coding performance and scalability options of state-of-the-art wavelet-based codecs [6].

The paper focuses on the optimality of embedded deadzone scalar-quantizers for high-rate $L_\infty$-constrained scalable wavelet-based coding of images. The embedded quantizers are derived by constructing a high-rate $L_\infty$-oriented rate-distortion (R-D) model, applicable to any family of embedded deadzone uniform scalar quantizers [7]. The model is validated experimentally and conclusions are drawn regarding the optimal subband quantizer instantiations [7]. While state-of-the-art wavelet-based $L_2$-oriented coders commonly employ embedded scalar quantizers having a double-sized deadzone [6], it is shown that this embedded quantizer instantiation should not be de facto employed in the $L_\infty$ case. Finally, the paper provides experimental results and the performance comparison of the coding scheme of [4, 5], equipped with the optimal quantizers, versus an embedded codec belonging to the same wavelet-based family [6] and $L_\infty$-oriented codecs that employ a different approach [2, 3].

II. OPTIMAL EMBEDDED $L$-INFINITE-CONSTRAINED SCALAR QUANTIZATION

A. Embedded Scalar Quantizer Output Entropy

We consider the family of embedded deadzone uniform scalar quantizers $\{Q_{\xi, n}\}$ in which every source sample $X$ is quantized to [6]:

$$q_{\xi, n} = Q_{\xi, n}(X) = \begin{cases} \text{sign}(X) \cdot \left(\frac{|X| + \xi}{2^n \Delta} \right) & \text{if } \frac{|X|}{2^n \Delta} + \frac{\xi}{2^n} > 0 \\ 0 & \text{otherwise} \end{cases}$$

(1)

where $\Delta > 0$, $\left\lfloor a \right\rfloor$ denotes the integer part of $a$, the parameter $\xi$ determines the width of the deadzone, and $n \in \mathbb{Z}_+$ represents the number of discarded bit-planes.
We restrict ourselves to mid-tread quantizers and the range of interest $\xi \in (-\infty, 1/2]$, i.e. embedded deadzone quantizers possessing a deadzone bin-size that is larger or equal to the other bin-sizes [6].

**Theorem 1:** Let $X$ be a memoryless source of randomly generated numbers whose density function can be modeled as a “generalized Laplacian” [6]. Then the output entropy $H(Q_{\xi,n}(X))$ of any quantizer $Q_{\xi,n}$ is given by [7]:

$$H(Q_{\xi,n}(X)) \approx h(X) - \log_2 \left[ 2^{n+1} \left( 1 - \frac{\xi}{2^n} \right) \Delta \right] + \log_2 \left[ 2 \left( 1 - \frac{\xi}{2^n} \right) \right] Q(p,z)$$  \hspace{1cm} (2)

where $h(X) = \log_2 \left( 2a^{-1} e^{\phi a} \sigma \cdot \Gamma \left( 1/a \right) \sqrt{\Gamma \left( 1/a \right) / \Gamma \left( 3/a \right)} \right)$. $\Gamma$ is the Gamma function, $\sigma$ is the standard deviation, $a$ is the density-function parameter [6], and $Q(p,z)$ is the regularized Gamma function, with $p = 1/a$, $z = \left( 2^n \sigma^{-1} \Delta \left( 1 - \frac{\xi}{2^n} \right) \cdot \sqrt{\Gamma \left( 3/a \right) / \Gamma \left( 1/a \right)} \right)^a$.

**Proof:** The proof is outlined in the Appendix.

**B. Embedded L-Infinite-Constrained Scalar Quantization Modeling**

We will refer throughout the paper to an arbitrary bitplane of a single subband (Subband BitPlane, SBP), respectively a complete bitplane (BitPlane, BP) covering all subbands.

It can be shown [4, 5] that for a lifting-based wavelet transform followed by uniform scalar quantization of the subbands, the MAXAD is given by:

$$M = \sum_{s=1}^{S} K_s \frac{\Delta_s}{2}$$  \hspace{1cm} (3)

where $S$ is the total number of subbands, $\Delta_s$ is the quantization bin size employed for subband $s$, $1 \leq s \leq S$, and $K_s$, $K_s \geq 0$ are the contribution factors [4, 5] derived from the prediction and update coefficients of the lifting scheme. The notations here refer to the subbands raster scanning
We apply embedded uniform deadzone quantizers to each subband \( s \), which we denote as \( \{Q^{(s)}_{\xi,n}\} \), where \( n, 0 \leq n \leq N \) indicates the quantization level (i.e. the BP) and \( N + 1 \) is the total number of BPs. Relation (3) can be formalized for the case of embedded uniform deadzone quantizers observing that for any \( Q^{(s)}_{\xi,n} \) with \( \xi \in (-\infty, 1/2] \), the largest bin-size is the size of the deadzone \([6]\). Thus, the expression of the MAXAD at the end of each SBP can be written as:

\[
M_{\xi,n,q} = \frac{q-1}{\mathcal{K}_{\xi} \cdot \Delta_{\xi,2^{n+1}} (1-\frac{\xi}{2^{n+1}}) + \sum_{s=q}^{S} \mathcal{K}_{s} \cdot \Delta_{s,2^{n}} (1-\frac{\xi}{2^{n}})}
\]

with \( 1 \leq q \leq S \). In the expression above, the first term corresponds to the MAXAD contribution of \( q-1 \) subbands having \( n+1 \) discarded SBPs, while the second term corresponds to the remaining \( S-q+1 \) subbands having \( n \) discarded SBPs. For \( q=1 \), relation (4) gives us the expression of the MAXAD after \( n \) discarded BPs.

The total rate \( R_{\xi,n,q} \) at the end of each SBP can be written as:

\[
R_{\xi,n,q} = \sum_{s=1}^{q-1} \beta_{s} R_{\xi,s}^{(s)} + \sum_{s=q}^{S} \beta_{s} R_{\xi,s}^{(s)}
\]

where \( \beta_{s} \) is the relative size of subband \( s \) with respect to the entire image and \( R_{\xi,s}^{(s)} \) is the subband rate derived using theorem 1. For \( q=1 \), relation (5) gives the rate after \( n \) discarded BPs.

For any \( \xi \), we optimize the highest-rate quantizers \( \{Q^{(s)}_{\xi,0}\}, 1 \leq s \leq S \), by solving the D-R optimization problem of minimizing the rate \( R_{\xi,0} \) given a MAXAD constraint \( M_{\xi,0} = M_{1} \). The solutions \( \{\Delta_{s}\} \) are computed using Lagrangian multipliers.
For $\xi = 1/2$ the highest-rate quantizers $\{Q^{(i)}_{l/2,0}\}$ are uniform quantizers [6]; it can easily be shown that the solutions of the optimization problem are given by $
abla_i = 2M_i \beta_i / K_i$. For any other $\xi$, i.e. $\xi < 1/2$, the Lagrangian formalism leads to transcendental equations that do not possess a set of analytical solutions. In general, for any $\xi \leq 1/2$, we propose the following analytical approximations of the solutions, which we validate numerically in Section III:

$$\nabla_i = M_i \beta_i / \left( K_i (1 - \xi) \right)$$

(6)

By replacing the solutions (6) in the expressions (4) and (5) for $q = 1$, we obtain the MAXAD and the total rate at the end of every BP [7]:

$$M_{\xi,n,1} = M_i 2^n \left( 1 - \xi / 2^n \right) / \left( 1 - \xi \right)$$

(7)

$$R_{\xi,n,1} = \sum_i \beta_i \left[ h(X_i) - \log_2 \left( M_i \beta_i 2^{x_i+1} \left( 1 - \xi / 2^n \right) / \left( K_i (1 - \xi) \right) \right) + \log_2 \left( 2 \left( 1 - \xi / 2^n \right) / \left( K_i (1 - \xi) \right) \right) Q(p_i, z_i) \right]$$

(8)

with $p_i = 1/a_i$ and $z_i = \left( 2^n \sigma_i^{-1} \left( 1 - \xi / 2^n \right) \cdot \left( M_i \beta_i / \left( K_i (1 - \xi) \right) \right) \cdot \sqrt{\Gamma(3/a_i) / \Gamma(1/a_i)} \right)^{x_i}$.

By combining the relations (7) and (8), the operational rate-distortion function (RDF), which we denote as the continuous function $R_{\xi}(M)$, can be written for any $\xi \leq 1/2$ as [7]:

$$R_{\xi}(M) = \sum_i \beta_i \left[ h(X_i) - \log_2 \left( \frac{2 \beta_i M}{K_i} \right) + \log_2 \left( \frac{2M \left( 1 - \xi \right)}{M \left( 1 - \xi \right) + M \xi} \right) Q \left( \frac{1}{a_i}, \frac{M \beta_i}{K_i}, \sqrt{\Gamma(3/a_i) / \Gamma(1/a_i)} \right) \right]$$

(9)

This function passes through all points of the form $(R_{\xi,n,1}, M_{\xi,n,1})$ and gives a continuous approximation of intermediate clusters of points $(R_{\xi,n,q}, M_{\xi,n,q})$.

The operational RDF $R_{\xi}(M)$ is written in (9) as a continuous function of the variable $M$, $M \geq M_i > 0$, given a fixed parameter $\xi$. The same expression can be rewritten as $R_{\xi}(M)$, i.e. as a function of the variable $\xi$, given a fixed $M$. 

5
Theorem 2. The continuous function $R_M(\xi)$ has the property that $\arg\min_{\xi \in (-\infty,\infty)}\left(R_M(\xi)\right) = 1/2$ for any $M$, $M \geq M_\gamma > 0$ [7].

Proof: The proof is outlined in the Appendix.

Based on theorem 2, we can derive for every wavelet subband $s$, that the optimal embedded-quantizer family instantiation is $\{Q_{(2,s)}^{(s)}\}$.

III. EXPERIMENTAL RESULTS

Model Validation

Table 1.a) illustrates the mean square error $MSE_\Delta$ between the proposed analytical solutions $\{\Delta_s\}$ given by (6), and the solutions obtained by using a numerical optimization technique; the results are obtained on the “Barbara” image for different values of $\xi$. The relative error between the highest-rate $R_{\xi,0}^{\text{model}}$ given by the model and the rate $R_{\xi,0}^{\text{exp}}$ obtained experimentally is also shown, together with the average relative errors between the experimental rates $R^{\text{exp}}$ and theoretical rates $R^{\text{model}}$ obtained for MAXAD values $M$, $M_i \leq M \leq M_T$, with $M_i, M_T$ chosen as $M_i = 1$ and $M_T = 35$. The average relative error over a few sample values of $\xi$, for $M_i \leq M \leq M_T$, has also been computed. Similar results are given in Table 1.b), over a sample set of ten 8-bit natural and remote sensing images, five of which where taken from the JPEG2000 test set [6].

The R-D curves of the model obtained on “Barbara” for $\{1/2, 0\}$ have been plotted in Fig. 1. The set of points on the model correspond to the guaranteed MAXAD values at the end of each SBP, i.e. as given by (4), and the corresponding rates given by (9); for these MAXAD values we also plot the rates obtained experimentally. The results shown in Table 1, together with those depicted in Fig. 1, confirm the validity of the proposed model. A set of model curves are plotted.
in Fig. 2.a), for different values of $\xi$; the areas contained underneath the segments obtained by interpolating the set of discrete experimental R-D points are plotted as cumulative functions in Fig. 2.b). The results illustrated in Fig. 2 confirm the theoretical conclusions of theorem 2.

**Coding Results**

Compression results are given in Table 2, for different target MAXAD values, comparing the results obtained by our Embedded Wavelet Maxad Coder (E-WMC) [4, 5] equipped with optimal quantizers against those obtained by the fixed $L_\infty$-oriented Lossy plus Near-Lossless (LNL) codec [3], the Progressive $L_\infty$-oriented Lossy/Near Lossless (PLNL) codec [2] and the progressive wavelet-based $L_2$-oriented JPEG2000 codec [6]; the results are reported for [2] in fixed $L_\infty$-coding mode. The obtained (actual) MAXAD is denoted in these tables as $\|e\|_\infty$; the guaranteed MAXAD, for the same bit-rate, is denoted as $\|g\|_\infty$. The overall MAXAD behaviors of the two embedded wavelet-based codecs E-WMC and JPEG2000 are illustrated in Fig. 3.

**IV. CONCLUSIONS**

In this paper we have analyzed the optimality of embedded deadzone scalar-quantizers for high-rate $L_\infty$-constrained scalable wavelet-based coding of images. We have derived and validated experimentally a rate-distortion model applicable to the family of embedded deadzone scalar-quantizers. Based on the model, conclusions have been drawn with respect to the optimal family instantiation for high-rate $L_\infty$ wavelet-based image coding. The optimal quantizers have been employed in a coding algorithm that, while outperformed by codecs that operate in the image domain and target fixed $L_\infty$-oriented compression, remains competitive with respect to similar embedded wavelet-based codecs that target $L_2$-oriented coding; moreover, it retains the capabilities of wavelet-based compression schemes, while providing a fully embedded $L_\infty$-constrained bit-stream.
APPENDIX

Proof Outline of Theorem 1: The output entropy of a high-resolution non-uniform scalar quantizer is given by [8]:

\[ H(Q(X)) = h(X) + \int f(y) \log_2 \Lambda(y) dy \]

where \( h(X) = -\int f(y) \log_2 f(y) dy \) is the differential entropy of the source variable \( X \) and \( \Lambda(y) \) is the unnormalized point density of the quantizer.

The unnormalized point densities \( \Lambda(y_{q,n}) \) of each quantizer \( Q_{q,n} \) can be written as

\[ \Lambda(y_{q,n} = 0) = 1/(2^n (1 - \xi/2^n) \Delta) \quad \text{and} \quad \Lambda(y_{q,n} = 0) = 1/(2^n \Delta). \]

By replacing these densities in relation (10), together with the expression \( f(y) \) of the generalized Laplacian density function, we obtain relation (2).

Proof Outline of Theorem 2: Starting from relation (9), the operational RDF can be written as:

\[ R_M(\xi) = b_M + c_M \cdot g_M(\xi) \]

where \( g_M(\xi) = \log_2 \left[ 2M (1 - \xi) / \left( M (1 - \xi) + M_i \xi \right) \right] \), and \( b_M, c_M \) are constant with respect to \( \xi \). It can easily be verified that \( \partial Q(p, z) / \partial z < 0 \) for any \( z > 0 \), given any fixed \( p > 0 \). Since \( Q(p, \infty) = 0 \), this implies that \( Q(p, z) > 0 \). Using this inequality in the expression of \( c_M \), derived from relations (9) and (11), we obtain that \( c_M > 0 \). In addition, it can easily be shown that \( g'_M(\xi) < 0 \) and \( g_M(\xi) \geq 0 \) for any \( \xi \in (-\infty, 1/2] \), implying that \( g_M(\xi) \) is a positive strictly decreasing function on this interval. We conclude that \( \arg \min_{\xi \in [-\xi, 1/2]} R_M(\xi) = 1/2 \) for any \( M, M \geq M_i > 0 \).
REFERENCES


<table>
<thead>
<tr>
<th>$\xi$</th>
<th>$\frac{1}{2}$</th>
<th>$\frac{1}{4}$</th>
<th>0</th>
<th>$-\frac{1}{4}$</th>
<th>$-\frac{1}{2}$</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE, $\xi$</td>
<td>$0.2 \times 10^{-4}$</td>
<td>$0.7 \times 10^{-3}$</td>
<td>$0.1 \times 10^{-3}$</td>
<td>$0.8 \times 10^{-5}$</td>
<td>$0.1 \times 10^{-5}$</td>
<td></td>
</tr>
<tr>
<td>$R_{\text{exp}}^\xi - R_{\text{mod}}^\xi$ / $R_{\text{mod}}^\xi$ (%)</td>
<td>1.82</td>
<td>1.46</td>
<td>1.28</td>
<td>1.18</td>
<td>1.12</td>
<td>2.39</td>
</tr>
<tr>
<td>$R_{\text{exp}}^\xi - R_{\text{mod}}^\xi$ / $R_{\text{mod}}^\xi$ (%)</td>
<td>3.25</td>
<td>1.77</td>
<td>3.02</td>
<td>2.15</td>
<td>1.73</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Model accuracy for a) “Barbara”, b) a test set of ten images, for different values of $\xi$.

<table>
<thead>
<tr>
<th>MAXAD</th>
<th>Lena</th>
<th>Barbara</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PLNL bpp</td>
<td>LNL bpp</td>
</tr>
<tr>
<td>1</td>
<td>2.58</td>
<td>2.78</td>
</tr>
<tr>
<td>3</td>
<td>1.50</td>
<td>1.95</td>
</tr>
<tr>
<td>7</td>
<td>0.73</td>
<td>1.40</td>
</tr>
</tbody>
</table>

Table 2. Comparison of bit-rates achieved using image-domain-based fixed $L_\infty$, wavelet-domain-based embedded $L_2$ and wavelet-domain-based embedded $L_\infty$ techniques.
Figure 1. Model and experimental R-D curves for $\xi = 1/2, 0$.

Figure 2. a) Model R-D curves for different parameter values $\xi$; b) Cumulated area plots of the experimental R-D points for different parameter values $\xi$.

Figure 3. Comparison on “Lena” of overall MAXAD behavior using wavelet-based embedded $L_2$ and $L_\infty$ techniques.