ℓ_2 Optimized Predictive Image Coding with ℓ_∞ Bound

Sceuchin Chuah, Sorina Dumitrescu, *Senior Member, IEEE,* and Xiaolin Wu, *Fellow, IEEE*

Abstract

In many scientific, medical and defense applications of image/video compression, an ℓ_∞ error bound is required. However, pure ℓ_∞-optimized image coding, colloquially known as near-lossless image coding, is prone to structured errors such as contours and speckles if the bit rate is not sufficiently high; moreover, most of the previous ℓ_∞-based image coding methods suffer from poor rate control. In contrast, the ℓ_2 error metric aims for average fidelity and hence preserves the subtlety of smooth waveforms better than the ℓ_∞ error metric and it offers fine granularity in rate control; but pure ℓ_2-based image coding methods (e.g., JPEG 2000) cannot bound individual errors as the ℓ_∞-based methods can. This paper presents a new compression approach to retain the benefits and circumvent the pitfalls of the two error metrics.

A common approach of near-lossless image coding is to embed into a DPCM prediction loop a uniform scalar quantizer of residual errors. The said uniform scalar quantizer is replaced, in the proposed new approach, by a set of context-based ℓ_2-optimized quantizers. The optimization criterion is to minimize a weighted sum of the ℓ_2 distortion and the entropy while maintaining a strict ℓ_∞ error bound. The resulting method obtains good rate-distortion performance in both ℓ_2 and ℓ_∞ metrics and also increases the rate granularity. Compared with JPEG 2000, the new method not only guarantees lower ℓ_∞ error for all bit rates, it even achieves higher PSNR for relatively high bit rates.

Index Terms

ℓ_∞-constrained image compression, predictive coding, optimal scalar quantization.

During the writing of this paper the first author was with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, L8S 4K1 Canada. The second and third authors are with the same department. E-mails: chuahs@mcmaster.ca, sorina@mail.ece.mcmaster.ca and xwu@mail.ece.mcmaster.ca.

I. INTRODUCTION

For many important applications of image compression in science, medicine, space exploration, precision engineering, etc., high fidelity of image reconstruction is required. The ideal solution would be mathematically lossless compression, but this can be done only with a high bit budget (currently achievable lossless rate is above 4 bpp for most natural images). A practical alternative is to accept some small, tightly controlled loss and trade for significantly reduced bit rates, which is referred to as near-lossless compression in the literature. A common criterion of near-lossless image compression is that a predefined upper bound of the reconstruction error is imposed on every single pixel; in this sense, near-lossless image compression is synonymous with ℓ_∞-constrained image compression.

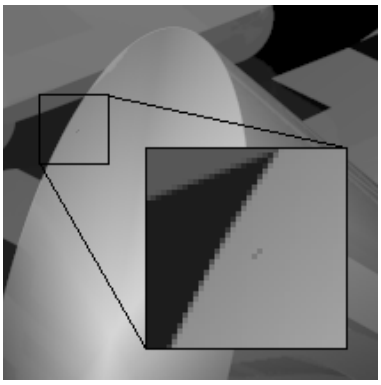
A straightforward method of ℓ_∞-constrained image compression is a cascade of uniform scalar quantization of all pixel values followed by lossless coding of the pre-quantized image [1], [2]. To achieve an ℓ_∞ bound τ, a step size of $2\tau + 1$ is used in the scalar quantizer. But a far more efficient near-lossless coding approach is a closed loop of a causal DPCM predictor and uniform scalar quantizer of prediction residuals [3]–[7]. The algorithm in [3] is a DPCM coding technique that employs context-based source modeling and arithmetic coding for lossless compression of quantized prediction errors. In order to achieve an ℓ_∞ bound of $\tau = 1$, two different scalar midtread quantizers were used. One of them is uniform with all step sizes equal to 3. The other one is nearly uniform with a quantizer bin of size 3 around 0 and all the other bins of size 2. The system in [4] is also based on the DPCM method, but it incorporates an additional mechanism to minimize the entropy of the sequence of quantized prediction residues using a so-called DPCM-trellis. The trellis state transitions restrict the possible pixel reconstructions to those within a τ-error bound. An iterative algorithm determines the trellis path corresponding to the minimum entropy sequence of quantized residuals for each image row. The ℓ_∞-constrained (or near-lossless) CALIC [5] is a variant of lossless CALIC [8], which incorporates a uniform scalar quantizer for the residual errors in the context-based prediction loop. In particular, a quantizer bin size of $(2\tau+1)$ is used to ensure no errors greater than τ.

Among the aforementioned ℓ_∞-constrained image coders, near-lossless CALIC achieves the highest compression performance when $\tau \leq 3$ [5]. Further enhancements of near-lossless CALIC were proposed in [6], [7] which led to superior performance in terms of bit rate and/or ℓ_2 distortion. However, these techniques have high computational complexity either at the encoder, in [6], where adaptive context modeling is used, or at the decoder, in [7], where the hard decision decoding is followed by an ℓ_2 image restoration step.

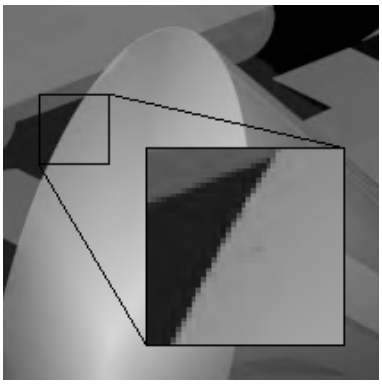
The design goal of the aforementioned ℓ_∞-constrained image coders is to achieve the lowest bit rate for each given error bound τ, neglecting other operational issues. One side effect is that the number of achievable bit rates is small, only equal to the number of possible values of τ. Such a coarse rate granularity makes it very difficult to finely adjust the bit rate versus the distortion bound. The problem is illustrated by Table I that lists the bit rates of the image in Fig. 7e achievable by near-lossless CALIC for $\tau = 0, 1, 2, \ldots, 8$ ($\tau = 0$ corresponds to lossless CALIC). Notice the big gaps between consecutive bit rates, especially as τ decreases, which is the case of interest. For instance, in order to improve the reconstruction quality at $\tau = 2$, the only option is to choose $\tau = 1$, but this incurs a big rate increment of 0.56 bpp (or 35.67%). The resulting next higher rate will be wasteful if the reconstruction quality at $\tau = 2$ is just slightly lower than required.

Moreover, the suitability of pure ℓ_∞ distortion metric in preserving image quality may also be put into question. In particular, ℓ_∞-constrained image coders may introduce structured artifacts in smooth regions even when the value of τ is as low as 3. Fig. 1 compares images coded by an ℓ_∞-based compression method (near-lossless CALIC) and an ℓ_2-based compression method (JPEG 2000) when the bit rates are the same. As shown in Fig. 1c, the ℓ_∞-constrained CALIC produces contours in the smooth shade region. Although the ℓ_2-based JPEG 2000 does not produce contour artifacts, the small feature on the smooth surface is barely noticeable, as identified in Fig. 1b. In fact, the tendency of ℓ_2-based image coders to distort or even remove small features, which are statistical outliers, motivated the research on ℓ_∞-based image coders. In some important applications, tiny objects (e.g., lesions in medical images or small boats in satellite images) carry great semantic significance even though they are tiny minority statistically speaking.

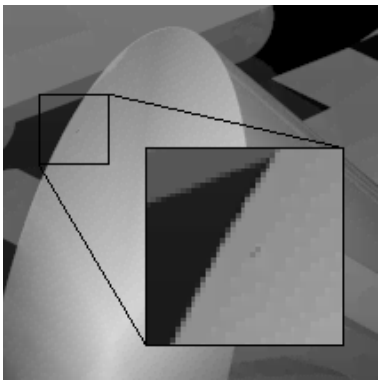
To summarize the above observations, the ℓ_2 distortion metric, being an average fidelity measure, preserves the subtle smooth image waveforms better; on the other hand, the ℓ_∞ distortion metric, aiming for best minmax approximation, preserves isolated small image features better. The other major difference between the ℓ_2 and ℓ_∞ code design criteria is that the former offers much finer rate granularity than the latter. Now a natural inquiry, which is the main theme of this paper, is in order: can one get the advantages of the two metrics but not their shortcomings? Towards this objective, we develop a new technique of incorporating a mechanism of ℓ_2 optimization in the existing framework of ℓ_∞-constrained image coding. Specifically, we modify the near-lossless CALIC system by replacing the previous in-loop uniform scalar quantizer with a set of context-based ℓ_2-optimized scalar quantizers. The main innovations of this work are the formulation of and an algorithm for the optimal code design problem of minimizing a weighted sum of the ℓ_2 distortion and the total rate over all possible quantizers, while obeying a specified ℓ_∞ error



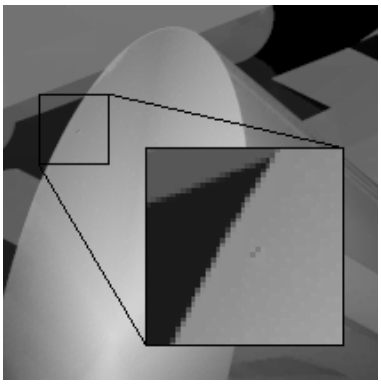
(a) Original image



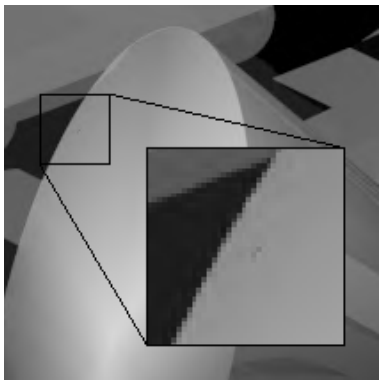
(b) JPEG 2000 at 0.45 bpp



(c) Near-lossless CALIC at 0.45 bpp



(d) Proposed at 0.51 bpp



(e) JPEG 2000 at 0.51 bpp

Fig. 1. Comparison using a computer-generated test image between (a) Original image / Lossless image coding (Lossless CALIC at rate 1.27 bpp); (b) ℓ_2-based image coding (JPEG 2000 at rate 0.45 bpp, PSNR 43.62 dB, ℓ_∞ error bound 21) with the small feature in the smooth region blurred; (c) ℓ_∞-based image coding (Near-lossless CALIC at rate 0.45 bpp, PSNR 41.58 dB, ℓ_∞ error bound 3) with speckles and contours as artifacts; (d) Proposed method (at rate 0.51 bpp, PSNR 41.74 dB, ℓ_∞ error bound 3) preserving the small feature with minimal artifacts; (e) ℓ_2-based image coding (JPEG 2000 at rate 0.51 bpp, PSNR 44.81 dB, ℓ_∞ error bound 17) with the small feature in the smooth region blurred.

bound.

Our work is closely connected in scope to [9]. The authors of [9] also aim at exploiting the advantages of both ℓ_2 and ℓ_∞ metrics. Their solution is a wavelet-based ℓ_∞-oriented scalable image coder. Their approach is to ensure that an upper bound on the maximum error in the image domain, which is expressed in terms of the transform-domain subband quantization bin sizes, satisfies the ℓ_∞ constraint. However, as the results reported in [9] show, the ℓ_∞ performance of this encoder is far inferior to that of near-lossless CALIC. In fact, it is well known that transform coding is inferior to predictive coding when the bit rate is sufficiently high. This is because for high rates, bits will be allocated to code small high frequency

TABLE I

ACHIEVABLE BIT RATES AND PSNR FOR IMAGE IN FIG. 7E USING NEAR-LOSSLESS CALIC

τ	Bit Rate (bpp)	PSNR (dB)
0	3.53	∞
1	2.13	49.95
2	1.57	45.38
3	1.27	42.53
4	1.07	40.47
5	0.94	38.84
6	0.82	37.43
7	0.73	36.30
8	0.67	35.25

coefficients in the transform domain. The energy packing advantage of transform coding will be lost. At lossless rate (the highest for a given source), all codecs have to reconstruct noises; coding in transform domain is clearly a poor choice in this case.

This paper is structured as follows. In the following section we briefly review the prior work on optimal (predictive loop) quantizer design and emphasize the contribution of our work versus previous designs. In Section III we briefly describe how near-lossless CALIC operates. Then in Section IV, we formulate the problem of minimizing the weighted sum of the ℓ_2 distortion and of the average entropy over all different context-based quantizers, under a common ℓ_∞ constraint. Further, we describe the solution algorithm based on the graph approach. We subsequently integrate the optimized quantizers obtained in Section IV into near-lossless CALIC to obtain the proposed image coder and present extensive experimental results in Section V. The results include performance comparisons with JPEG 2000 and near-lossless CALIC. Finally, conclusions are given in Section VI.

II. RELATION TO PRIOR WORK ON OPTIMAL (PREDICTIVE LOOP) QUANTIZER DESIGN

Optimization of the quantizer used in the prediction loop has been proposed in the past in [10]–[12] for the scalar case and in [13]–[15] for the vector case. One difficulty encountered when addressing this problem resides in obtaining a distribution that accurately represents the distribution of the prediction errors. This is because the statistics of the residuals depends on the quantizer. Most authors have generated the training set of prediction errors by using the unquantized pixel values in the prediction. This method is known as the open-loop (OL) approach. To address the statistical mismatch of OL, the closed-loop

(CL) and the asymptotic closed-loop (ACL) approaches were proposed in [13] and [14], respectively. They performed the design process iteratively, the quantizer optimized at each iteration being used to obtain the training set of residuals for the next iteration.

In this work we adopt the OL approach. The reasoning behind this choice is that, in near-lossless compression small values of τ are of interest, and in this case, the OL approach provides a good enough approximation of the true statistics of prediction errors. We point out that in this work, after collecting the statistics of residuals for each context from a training set of images, each conditional distribution is approximated by a Laplacian distribution which is further used in the optimization.

What distinguishes our work from previous work on optimal quantizer design is mainly the criterion used in the optimization. Most quantizer design algorithms aim at minimizing the ℓ_2 distortion for fixed number of quantizer levels, or minimizing a weighted sum of the ℓ_2 distortion and the entropy. We are not aware of any work which incorporates the ℓ_∞ constraint alongside. Scalar quantizer design algorithms mainly fall into one of the following categories: 1) Lloyd-Max method [10], [16], which iteratively optimizes the encoder and the decoder, respectively, while keeping the other component fixed, and 2) combinatorial algorithms [17]–[24]. While the first approach ensures only a locally optimal solution, the latter algorithms guarantee global optimality when the source alphabet is finite. Our optimization problem requires simultaneous optimization of all quantizers corresponding to different contexts under a common constraint on the ℓ_∞ error bound. Interestingly, the separability of the cost function allows for separate optimization of each quantizer. We further show that the latter problem can be modeled as a minimum weight path problem. This model is similar in spirit to that used in [24]. However, we emphasize that while in [24] only the minimization of the weighted sum of the ℓ_2 distortion and entropy was considered, our problem is different due to the additional ℓ_∞ constraint.

The proposed image coder is able to achieve a much denser set of bit rates than near-lossless CALIC. As our experiments performed on images outside the training set show, the ℓ_∞ constraint enforced in our algorithm allows us to achieve ℓ_∞ error bounds that are always lower than those of JPEG 2000. Meanwhile, the minimization of the ℓ_2 distortion incorporated in the design leads to better ℓ_2 performance than JPEG 2000 above a certain threshold bit rate for each image, threshold which can be as low as 1.1 bpp. Additionally, the fine granularity allows for the reconstruction quality to be improved by adding only small amounts to the used bit rate. In particular, as it can be seen in Fig. 1d the proposed coder eliminates the artifacts observed in Figs. 1b and 1c respectively, at the expense of only a very small

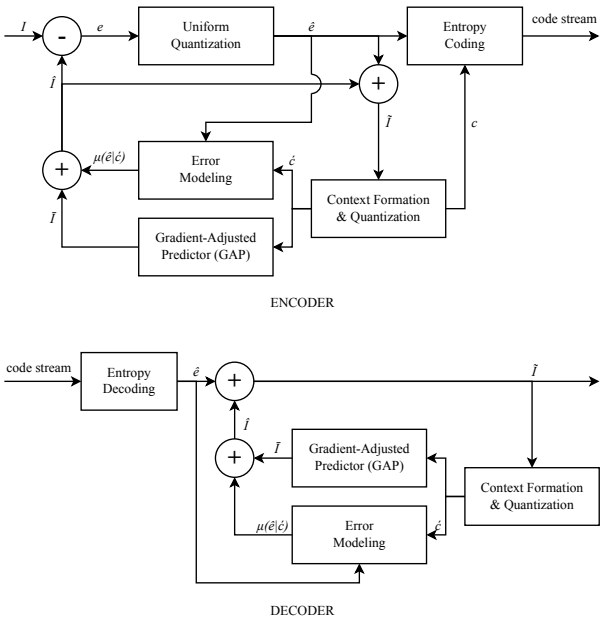


Fig. 2. Schematic description of near-lossless CALIC.

increase in the bit rate[1]. On the other hand, the same rate increase with the JPEG 2000 encoder does not achieve the same visual qulity as the proposed method, as illustrated by Fig. 1e.

III. NEAR-LOSSLESS CALIC

As shown in the flow diagram in Fig. 2, l_∞-constrained CALIC in [5] consists of five main components: gradient-adjusted prediction (GAP), uniform quantization, context formation and quantization, context modeling, and entropy coding.

We will only briefly describe the encoder since the decoder is just the encoder process reversed. Let I be the current pixel value to be encoded. The GAP module makes a prediction $\bar{I}$ of I based on the knowledge of the reconstructed pixels in a precisely defined neighbourhood. For this, an estimate of the gradient of the intensity function at the current pixel is made to guide the construction of $\bar{I}$. The prediction $\bar{I}$ is further improved to $\hat{I}$ by adding the conditional sample mean of the quantized prediction errors $\mu(\hat{e}|\acute{c})$ conditioned on the error modeling context $\acute{c}$. The number of error modeling contexts $\acute{c}$ considered in CALIC is 576 or higher and they are formed based on both the energy level and image

[1]Notice that the next achievable rate with near-lossless CALIC, corresponding to $\tau = 2$, is 0.56 bpp, hence higher than the rate 0.51 used in the proposed method.

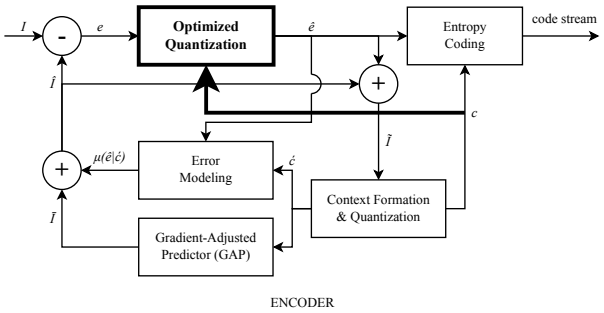


Fig. 3. Schematic description of the proposed image coder.

texture. The resulting prediction error (or residue) $e = I - \hat{I}$ is then quantized with a uniform scalar quantizer generating $\hat{e}$. The reconstructed pixel value $\tilde{I} = \hat{I} + \hat{e}$ and the quantized prediction residue $\hat{e}$ are fed back into the system to be used in the encoding of future pixels in the image.

Finally, the sequence of quantized prediction residues is losslessly entropy coded by means of a context-based arithmetic coder. Only eight coding contexts c are used for this purpose, which are formed by quantizing an error energy estimator Δ into eight bins. More specifically, Δ is the weighted sum of gradients of the reconstructed pixel values in the neighbourhood.

IV. Optimal Context-based Quantization of Prediction Errors

The proposed image coder shown in Fig. 3 replaces the uniform quantizer of prediction errors in near-lossless CALIC by scalar quantizers optimized for each individual coding context. The optimization problem aims at minimizing a weighted sum of the average ℓ_2 distortion and of the average entropy over all eight coding contexts, while preserving a maximum error bound for each prediction error. The optimization is performed assuming known distributions of prediction errors and coding contexts.

In this section we will address this optimization problem. Subsection IV-A presents the definitions and notations pertinent to ℓ_∞-constrained scalar quantizers. Subsection IV-B introduces the mathematical statement of the problem, which turns out to be solvable by separately optimizing the quantizer for each coding context. The following subsection presents the solution to the latter problem by modeling it as a minimum weight path problem in a weighted directed acyclic graph. Finally, the last subsection analyzes the computational effort.

A. ℓ_∞-constrained Scalar Quantizer

A quantizer maps the source alphabet into a smaller set of reproduction values. In our case, the source alphabet is a finite set of prediction residues $\mathcal{E} = \{e_n\}_{n=1}^{N}$, where $\{e_1 < e_2 < \cdots < e_N\}$. For raw input images using B bits per pixel, one has $N = 2^{B+1} - 1$ and $e_i = -2^B + i$, for all $1 \leq i \leq N$.

The encoder of a scalar quantizer is described by the partition that segments the source alphabet into a set of non-overlapping contiguous codecells. In other words, the encoder partition $\mathcal{P}$ can be defined as $\mathcal{P} = \{\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_K\}$ for some $1 \leq K \leq N$, where

$$\mathcal{C}_i = (a_{i-1}, a_i] = \{e_n \mid a_{i-1} < n \leq a_i\} \tag{1}$$

with $a_0 = 0$, $a_K = N$ and $0 \leq a_{i-1} < a_i \leq N$ for all $1 \leq i \leq K$.

The decoder of the quantizer, on the other hand, is described by the set of reproduction codewords $\{y_i \mid 1 \leq i \leq K\}$. Every alphabet symbol in codecell $\mathcal{C}_i$ is mapped to the reproduction codeword y_i. In an ℓ_∞-constrained quantizer with a maximum error bound of τ per symbol, the reproduction codeword must satisfy the condition $|e_n - y_i| \leq \tau$, for all $e_n \in \mathcal{C}_i$. Additionally, recall that our optimization criterion aims at minimizing a weighted sum of the ℓ_2 distortion and of the entropy. Since the choice of y_i does not affect the entropy of the quantized output, minimizing the aforementioned weighted sum reduces to minimizing the ℓ_2 distortion. Therefore, we determine y_i as follows

$$y_i = \underset{y \in \mathcal{E}, |e_n - y| \leq \tau, e_n \in \mathcal{C}_i}{\arg\min} \sum_{e_n \in \mathcal{C}_i} p(e_n)(e_n - y)^2, \tag{2}$$

where $p(e_n)$ is the probability of symbol e_n.

It is known that without the constraints $y \in \mathcal{E}$ and $|e_n - y| \leq \tau$ for all $e_n \in \mathcal{C}_i$, the value y_i minimizing the cost function in (2) is equal to the centroid of the codecell [10], [16], i.e., to

$$\mu(\mathcal{C}_i) = \sum_{e_n \in \mathcal{C}_i} p(e_n) \frac{e_n}{p(\mathcal{C}_i)}, \tag{3}$$

where $p(\mathcal{C}_i) = \sum_{e_n \in \mathcal{C}_i} p(e_n)$.

Now notice that due to the ℓ_∞-constraint, for the solution to (2) to exist, the size of all codecells must be limited to at most $(2\tau + 1)$. This means that for each $\mathcal{C}_i = (a_{i-1}, a_i]$, the condition $(a_i - a_{i-1}) \leq (2\tau + 1)$ has to be satisfied. Additionally, to achieve the ℓ_∞-constraint requirement, we must also ensure that the reproduction codeword y_i is at distance at most τ from the values at the boundaries of the codecell $\mathcal{C}_i$. Notice that the objective function in (2) is a quadratic function of y that is symmetrical around its point of minimum, i.e., around $\mu(\mathcal{C}_i)$. Therefore, the optimal solution to (2) is the point in $\mathcal{E}$ within distance

τ from the boundaries of the codecell that is closest to $\mu(\mathcal{C}_i)$, i.e.,

$$y_i = \begin{cases} (e_{a_{i-1}} + 1) + \tau, & \text{if } (\mu(\mathcal{C}_i) - (e_{a_{i-1}} + 1)) > \tau \\ e_{a_i} - \tau, & \text{if } (e_{a_i} - \mu(\mathcal{C}_i)) > \tau \\ [\mu(\mathcal{C}_i)], & \text{otherwise.} \end{cases} \tag{4}$$

where $[\mu(\mathcal{C}_i)]$ denotes the closest integer to $\mu(\mathcal{C}_i)$.

B. Optimization Problem Formulation

By optimizing the reproduction codewords for each encoder partition via (4), the ℓ_2 distortion and the output entropy corresponding to a quantizer become only functions of the encoding partition. Let us denote the ℓ_2 distortion and the output entropy for each codecell $\mathcal{C}_i$ as

$$d(\mathcal{C}_i) = \sum_{e_n \in \mathcal{C}_i} p(e_n)(e_n - y_i)^2, \;\; r(\mathcal{C}_i) = -p(\mathcal{C}_i) \log_2 p(\mathcal{C}_i) \tag{5}$$

respectively. Then the ℓ_2 distortion and the output entropy corresponding to a quantizer with encoder partition $\mathcal{P}$ are

$$D(\mathcal{P}) = \sum_{\mathcal{C} \in \mathcal{P}} d(\mathcal{C}), \;\; R(\mathcal{P}) = \sum_{\mathcal{C} \in \mathcal{P}} r(\mathcal{C}), \tag{6}$$

respectively. Now let us denote by $\mathcal{P}_m$ the encoder partition corresponding to the scalar quantizer for coding context c_m, where $1 \leq m \leq M$ and $M = 8$ for near-lossless CALIC. Subsequently, let D_T and R_T, respectively, denote the expected ℓ_2 distortion and entropy over the quantizers for all M contexts as follows

$$D_T = \sum_{m=1}^{M} q(c_m) D(\mathcal{P}_m), \;\; R_T = \sum_{m=1}^{M} q(c_m) R(\mathcal{P}_m), \tag{7}$$

where $q(c_m)$ is the probability of context c_m It is important to note that in the computation of $D(\mathcal{P}_m)$ and $R(\mathcal{P}_m)$ using (6) and (5) the probability $p(e_n)$ has to be replaced by the conditional probability of residual e_n conditioned on context c_m.

After having established the above notations we can now formulate the optimization problem as follows

$$\min_{\{\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_M\}} \;\; D_T + \gamma R_T, \tag{8}$$

for some $\gamma > 0$, where the optimization is performed over all possible M-tuples of partitions $\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_M$ with codecells of maximum size $2\tau + 1$.

We emphasize that using the weighted sum of the ℓ_2 distortion and of the entropy as the cost function is very common in the optimal entropy-constrained quantizer design [14], [24], [25]. The solution to

problem (8) corresponds to a point $({R_T}^*, {D_T}^*)$ on the lower convex hull of the set $\mathcal{RD}$ of all possible planar points of coordinates (R_T, D_T), such that the slope of a tangent to the set $\mathcal{RD}$ drawn through $({R_T}^*, {D_T}^*)$ is equal to $-\gamma$ [26]. Therefore, as γ decreases towards 0, the value ${R_T}^*$ is non-decreasing, while ${D_T}^*$ is non-increasing.

Substituting (7) back into (8) and rearranging, it follows that (8) is equivalent to

$$\sum_{m=1}^{M} \left\{ q(c_m) \min_{\mathcal{P}_m} J(\mathcal{P}_m, \gamma) \right\} \tag{9}$$

where $J(\mathcal{P}_m, \gamma) = D(\mathcal{P}_m) + \gamma R(\mathcal{P}_m)$. This shows that we can minimize the cost function (8) by individually minimizing $J(\mathcal{P}_m, \gamma)$ for each context c_m. Note also that since the ℓ_2 distortion $D(\mathcal{P}_m)$ and rate $R(\mathcal{P}_m)$ are additive over codecells, the cost $J(\mathcal{P}_m, \gamma)$ is also additive over codecells, in other words, the following holds

$$J(\mathcal{P}_m, \gamma) = \sum_{\mathcal{C} \in \mathcal{P}_m} j(\mathcal{C}, \gamma) \tag{10}$$

where $j(\mathcal{C}, \gamma) = d(\mathcal{C}) + \gamma r(\mathcal{C})$.

C. Solution Using the Minimum Weight Path Model

Due to the additive nature of the cost shown in (10), the task of minimizing (10) can simply be viewed as a single-source minimum weight path problem in a weighted directed acyclic graph (WDAG). More precisely, the set of vertices of the WDAG is $V = \{0, 1, \ldots, N\}$ and the set of edges is $E = \{(x, y) | x, y \in V, 0 < y - x \leq 2\tau + 1\}$ An edge (x, y) symbolizes a possible codecell $\mathcal{C} = (x, y]$ and its weight is defined as $w(x, y) = j(\mathcal{C}, \gamma)$.

A path in the graph is a sequence of connected edges and the weight of a path is the sum of the weights of all edges which make up that path. It is clear that any path in the graph from 0 to N is in unique correspondence with a partition $\mathcal{P}_m$. Furthermore, from (10), we see that the Lagrangian cost $J(\mathcal{P}_m, \gamma)$ of the partition equals the weight of the path. Hence, a path signifies a partition, the weight of a path equals the cost of the partition the path signifies, and minimizing the weight of a path is equivalent to minimizing the cost of the partition in (10) that corresponds to that path[2].

[2]Notice that this graph model is very close to the graph model used in [24] for the ℓ_2 optimization of an entropy constrained scalar quantizer. The difference between the two models stems from the fact that in [24] the ℓ_∞ constrained is not imposed. Therefore the graph in [24] contains every pair (x, y) with $0 \leq x < y \leq N$, as an edge. Additionally, the codeword y_i used in the computation of the weight of the edge, is not constrained to be within distance τ from the codecell boundaries.

(a) 1800×1800 pixels (b) 1701×1701 pixels (c) 1700×1700 pixels (d) 1601×1601 pixels

Fig. 4. Training set images.

Let $W(x, y]$ be the weight of the minimum weight path from x to y for $0 \leq x < y \leq N$. Then the problem solution is the path achieving $W(0, N]$. To determine $W(0, N]$ we compute all minimum weights $W(0, z]$ in increasing order of z, from 1 to N, by using the following recurrence

$$W(0, z] = \min_{y \geq 0, z-(2\tau+1) \leq y < z} \{W(0, y] + w(y, z]\}, \tag{11}$$

where $W(0, 0] = 0$.

D. Computational Cost

In order to solve the minimum weight path problem in the WDAG a preprocessing step which computes the edge weights is required. Since the number of edges is $O(\tau N)$ and computing the weight of an edge takes $O(\tau)$ time, the total number of operations needed in the preprocessing step is $O(\tau^2 N)$. Further, solving (11) for all z takes $O(\tau N)$ operations. Summarizing, the running time to minimize (10) is $O(\tau^2 N)$. Accounting for all M contexts, the running time to solve problem (8) becomes $O(\tau^2 M N)$. Note that in practice M is a small constant ($M = 8$) and the values of interest for τ are small as well, thus we may assume that τ is upper bounded by a constant. It follows that the time complexity to solve (8) is $O(N)$.

V. Experimental Results

A training set of four 8-bit high resolution images, shown in Fig. 4, were used to obtain the probability distributions of prediction errors for every context c_m and every value of $\tau \in \{1, 2, \cdots, 8\}$. We point out that the distributions corresponding to different values of τ are generally different since the contexts are different. Further, each of those distributions was approximated by a Laplacian distribution centered at zero. The approximations were done by choosing the Laplacian probability mass functions (pmfs) with

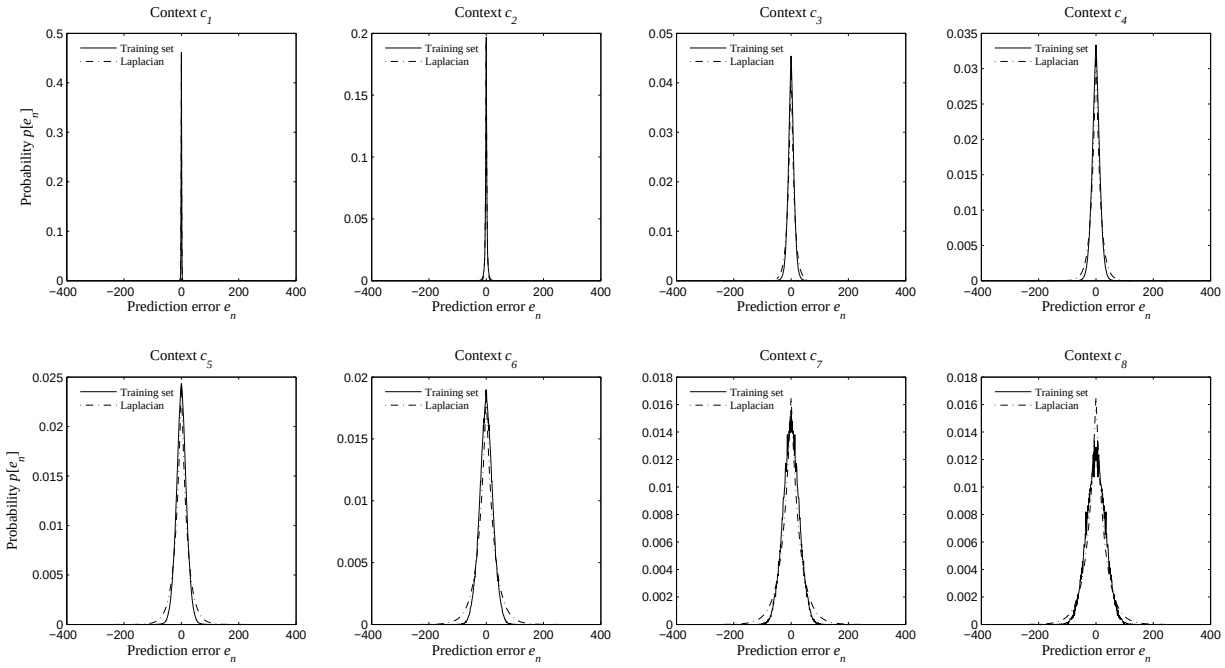


Fig. 5. Laplacian distributions approximating the distributions of prediction errors from the training set for $\tau = 5$.

the smallest average difference from the actual pmfs of the training set, i.e., by solving

$$\min_{b>0} \frac{1}{N} \sum_{n=1}^{N} \left(p_{lap}(e_n, b) - p_t(e_n) \right),$$

where

$$p_{lap}(e_n, b) = \begin{cases} \frac{1}{T}\left(1 - e^{-\frac{1}{2b}}\right) & \text{if } e_n = 0 \\ \frac{1}{T}\left(e^{-\frac{|e_n|-0.5}{2b}} - e^{-\frac{|e_n|+0.5}{2b}}\right), & \text{otherwise,} \end{cases}$$

and $T = 1 - e^{-\frac{N}{2b}}$, while $p_t(e_n)$ denotes the probability collected from the training set. An example of the approximations for $\tau = 5$ are shown in Fig. 5, and it can be seen that the Laplacian distributions very closely approximate the distributions of prediction errors. The approximations are necessary to obtain more generalized distributions which take into account random or outlying residue values not found in the training set.

For each $\tau \in \{1, 2, \cdots, 8\}$, we have solved (8) for a decreasing sequence of values of γ, starting with some high value $\gamma_{0,\tau}$. For each τ the partitions $\mathcal{P}_1^*, \mathcal{P}_2^*, \ldots, \mathcal{P}_M^*$, corresponding to the optimal solution for $\gamma_{0,\tau}$, are very close or even identical to the uniform quantizers in near-lossless CALIC, which have the largest possible step size of $(2\tau + 1)$. We will denote by $R_0(\tau)$ the value of R_T corresponding to this

solution. Then all the values R_T achieved for the same τ are larger than $R_0(\tau)$. Furthermore, one has $R_0(\tau+1) < R_0(\tau)$. Further, in order to proceed to testing the proposed coder on real images we have selected for each τ only those M-tuples for which R_T satisfies the condition $R_0(\tau) \leq R_T < R_0(\tau-1)$. The pairs $(R_T, PSNR_T)$ corresponding to these solutions are plotted in Fig. 6, where $PSNR_T = 10\log_{10}\frac{255^2}{D_T}$. The test images used in our simulations are shown in Fig. 7. They were chosen such that to cover a wide range of textures. We point out that the last three images were cropped out of images from the training set.

Since the authors of [6] had already experimented on incorporating the DPCM trellis implemented in [4] into near-lossless CALIC, and concluded that it did not offer appreciable compression gains despite the high computational complexity incurred, we will not further attempt to compare our work, which improves upon near-lossless CALIC, to the work in [4]. We will, however, make comparisons of our proposed solution with JPEG 2000, in terms of both ℓ_2 and ℓ_∞ performance. The parameters settings for the JPEG 2000 encoder are the following. The CDF $9/7$ wavelet transform is used. The number of decomposition levels is 4. The codeblock dimension is 64-by-64. The tile size equals the image size and the number of quality layers is 1.

Fig. 8 plots the ℓ_∞ error bound versus bit rate. It is clear from the figure that the proposed coder is significantly superior to JPEG2000 in terms of ℓ_∞ norm for all achievable bit rates. Fig. 9 plots the PSNR versus bitrate. As it can be seen the proposed coder outperforms JPEG2000 for rates higher than some image specific threshold, which can be as low as 1.1 bpp. Furthermore, Fig. 8 and 9 show that the proposed coder can achieve all 8 bit rates achievable by near-lossless CALIC with equal performance in terms of ℓ_∞ and ℓ_2 distortions. Additionally, the proposed coder achieves many intermediate bit rates, consistently improving the PSNR as the rate increases, while maintaining the same error bound as near lossless CALIC.

The proposed approach implies optimization of the quantizers, but performing it online increases the complexity of the encoder. Therefore, in order to keep the encoding complexity low, one option is to perform the optimization offline on a training set and store a number of such optimized M-tuples of quantizers at the encoder. In this case, the encoder needs to transmit to the decoder a label indicating which M-tuple is used, as side information; the resulting rate overhead is however negligible. The number of values of τ and rates covered can be chosen by taking into consideration particular requirements of the application. For each M-tuple of quantizers, the average entropy R_T achieved on the training set or the average bit rate obtained on coded images can be additionally stored in order to help estimate the achievable rate for a particular image. For a given M-tuple of quantizers, if there are some K_m number

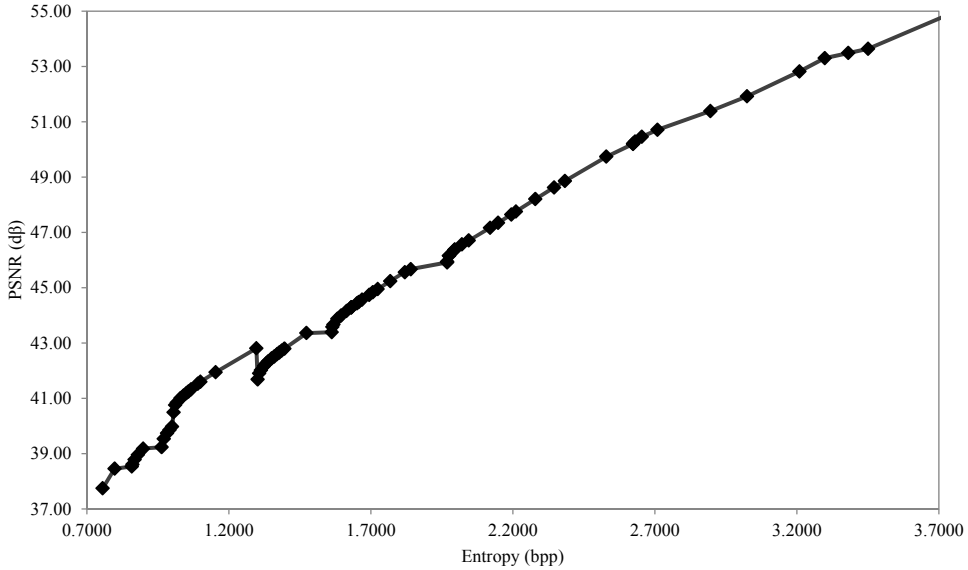


Fig. 6. PSNR versus average entropy R_T for the optimal M-tuples of quantizers used in our tests. The PSNR and the average entropy are computed using the distributions employed in the optimization.

of codecells for each context c_m, we only need to store the eight values of $\{K_m\}_{m=1}^8$, $\sum_{m=1}^8 K_m$ codewords and $\sum_{m=1}^8 K_m$ codecells. Thus, for an 8-bit per pixel input image, the memory needed amounts to $(16 + 2\sum_{m=1}^8 K_m)$ bytes in total, since the positive and negative values can be stored in separate tables. Further, since $K_m \leq N$ for every m, the total memory needed for storing an M-tuple of quantizers does not exceed 8kB. The storage requirement can be halved by forcing the quantizers to be symmetric. Note that in our tests we have imposed such a restriction and modified the algorithm for minimum weight path described in subsection IV-C accordingly. This restriction is motivated by the assumption that for the Laplacian distribution the optimal quantizer is symmetric or nearly so [27].

Out of the stored M-tuples of codebooks the user chooses one according to the specifications of the applications. This raises the question of how to accurately and effectively estimate the bit rate achieved for a particular image from the value R_T. Unfortunately, we do not have yet a low complexity solution to this problem, which is left for future work. On the other hand, if the specification on the target bit rate is rather a looser qualitative requirement, such as ”low”, ”medium” or ”high” bit rate, then selecting an appropriate M-tuple of quantizers could be easily done. For instance, if the specified ℓ_∞ constraint is τ and the used bit rate has to be low the the obvious choice is the M-tuple operating at the smallest rate,

(a) Hair (256×256 pixels)

(b) Plant (400×600 pixels)

(c) Flowers (256×256 pixels)

(d) Plants (256×256 pixels)

(e) Fruits (600×400 pixels)

Fig. 7. Test images.

i.e., with $R_T = R_0(\tau)$ or close to this rate. On the opposite, if high bit rate can be afforded, the M-tuple operating at rate $R_0(\tau - 1)$ can be chosen, while for moderate rate a value $R_T \approx (R_0(\tau) + R_0(\tau - 1))/2$ can be selected.

In some applications however, such as image archiving, the encoder can afford high complexity while decoding complexity still has to be low. In such cases the proposed technique can be used with an accurate rate control by trying several M-tuples of quantizers among the stored ones in a bisection search fashion until a rate close enough to the target rate is achieved. The online optimization with training set collected from the image at hand can also be incorporated at the encoder. The technique of [6] can be further applied with the potential of additional gain.

We emphasize that in the case when the optimization is performed online based on the image at hand, it is sufficient to transmit to the decoder only the coefficients defining the Laplacian distributions used in the optimization and the value of γ. Based on these coefficients the decoder performs the same

optimization as the encoder to determine the M-tuple of quantizers. This way the side information needed to be transmitted to the decoder is negligible.

VI. Conclusion

This research aims to improve the ℓ_2 performance and rate granularity of ℓ_∞-constrained image coding techniques. The traditional ℓ_∞-constrained coding method of uniformly quantizing prediction residuals is replaced by a set of context-based ℓ_2-optimized quantizers. The quantizer design criterion is to minimize a weighted sum of the ℓ_2 distortion and the entropy while imposing an upper bound on the quantizer cell size. The proposed technique obtains a good balance between ℓ_∞ and ℓ_2 performances. It ensures tighter ℓ_∞ error bound than JPEG 2000 for all bit rates, and at the same time it is competitive in ℓ_2 performance as well.

References

[1] A. Zandi, J.D. Allen, E.L. Schwartz and M. Boliek, "CREW: Compression with Reversible Embedded Wavelets," *Proc. Data Compression Conf. (DCC)*, 1995, vol., no., pp.212-221, 28-30 Mar. 1995.

[2] A. Said and W.A. Pearlman,"An image multiresolution representation for lossless and lossy compression," *IEEE Trans. on Image Process.*, vol.5, no.9, pp.1303-1310, Sep. 1996.

[3] K. Chen and T.V. Ramabadran, "Near-lossless compression of medical images through entropy-coded DPCM," *IEEE Trans. on Medical Imaging*, vol.13, no.3, pp.538-548, Sep. 1994.

[4] L. Ke and M.W. Marcellin, "Near-lossless image compression: minimum-entropy, constrained-error DPCM," *IEEE Trans. on Image Process.*, vol.7, no.2, pp.225-228, Feb. 1998.

[5] *A Context-Based, Adaptive, Lossless/Nearly-Lossless Coding Scheme for Continuous-Tone Images*, ISO/IEC Standard JTC 1.29.12, 1995.

[6] X. Wu and P. Bao, "L_∞ constrained high-fidelity image compression via adaptive context modeling," *IEEE Trans. on Image Process.*, vol.9, no.4, pp.536-542, Apr. 2000.

[7] X. Wu, J. Zhou and H. Wang, "High-Fidelity Image Compression for High-Throughput and Energy-Efficient Cameras," *Data Compression Conf. (DCC)*, 2011, vol., no., pp.433-442, 29-31 Mar. 2011.

[8] X. Wu and N. Memon, "Context-based, adaptive, lossless image coding," *IEEE Trans. on Commun.*, vol.45, no.4, pp.437-444, Apr. 1997.

[9] A. Alecu, A. Munteanu, J. P. H. Cornelis, and P. Schelkens, "Wavelet-based scalable L-infinity-oriented compression", *IEEE Trans. Image Proc.*, vol. 15, no. 9, pp. 2499–2512, Sept. 2006.

[10] J. Max, "Quantizing for minimum distortion," *IRE Trans. on Inform. Theory*, vol.6, no.1, pp.7-12, Mar. 1960.

[11] A. N. Netravali, "On Quantizers for DPCM Coding of Picture Signals," *IEEE Trans. on Inform. Theory*, vol. IT-23, no. 3, pp. 360-370, May 1977.

[12] D. K. Sharma and A. N. Netravali, "Design of Quantizers for DPCM Coding of Picture Signals," *IEEE Trans. on Commun.*, vol. COM-25, no. 11, pp. 1267-1274, Nov. 1977.

[13] V. Cuperman and A. Gersho, "Vector Predictive Coding of Speech at 16 kbits/s," *IEEE Trans. on Commun.*, vol. COM-33, no. 7, pp. 685-696, Jul. 1985.

[14] H. Khalil, K. Rose and S. L. Regunathan, "The Asymptotic Closed-Loop Approach to Predictive Vector Quantizer Design with Application in Video Coding," *IEEE Trans. on Image Process.*, vol. 10, no. 1, pp. 15-23, Jan. 2001.

[15] H. Khalil and K. Rose, "Predictive Vector Quantizer Design Using Deterministic Annealing," *IEEE Trans. on Signal Process.*, vol. 51, no. 1, pp. 244-254, Jan. 2003.

[16] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. on Inform. Theory*, vol.28, no.2, pp. 129- 137, Mar. 1982.

[17] J. D. Bruce, "Optimum quantization," Sc. D. thesis, M. I. T., May 14, 1964.

[18] D. K. Sharma, "Design of absolutely optimal quantizers for a wide class of distortion measures," *IEEE Trans. on Inform. Theory*, vol. IT-24, pp. 693-702, Nov. 1978.

[19] X. Wu, "Optimal Quantization by Matrix Searching," *J. of Algorithms*, 12(1991), vol.12, no. 4, pp. 663-673, Dec. 1991.

[20] X. Wu and K. Zhang, "Quantizer monotonicities and globally optimal scalar quantizer design," *IEEE Trans. on Inform. Theory*, vol. 39, pp. 1049-1053, May 1993.

[21] S. Dumitrescu and X. Wu, "Algorithms for optimal multi-resolution quantization," *J. Algorithms*, 50(2004), vol. 50, no. 1, pp. 1-22, Jan. 2004.

[22] S. Dumitrescu and X. Wu, "Optimal two-description scalar quantizer design," *Algorithmica*, vol. 41, no. 4, pp. 300. 269-287, Feb. 2005.

[23] S. Dumitrescu and X. Wu, "Lagrangian Optimization of Two-description Scalar Quantizers," *IEEE Trans. on Inform. Theory*, vol. 53, no. 11, pp. 3990-4012, Nov. 2007.

[24] D. Muresan and M. Effros, "Quantization as Histogram Segmentation: Optimal Scalar Quantizer Design in Network Systems," *IEEE Trans. on Inform. Theory*, vol.54, no.1, pp.344-366, Jan. 2008.

[25] P.A. Chou, T. Lookabaugh and R.M. Gray, "Entropy-constrained vector quantization," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol.37, no.1, pp.31-42, Jan. 1989.

[26] D. G. Luenberg, *Optimization by Vector Space Methods*, John Wiley & Sons, New York, 1969.

[27] G.J. Sullivan, "Efficient scalar quantization of exponential and Laplacian random variables," *IEEE Trans. on Inform. Theory*, vol.42, no.5, pp.1365-1374, Sep. 1996.

PLACE
PHOTO
HERE

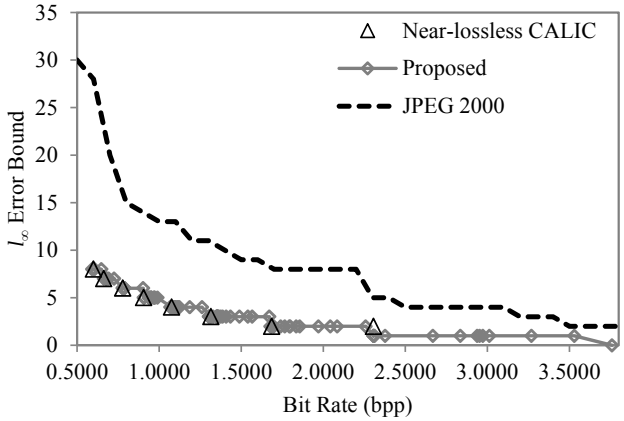
Sceuchin Chuah received her B.Eng. degree in electrical engineering and M.A.Sc. degree in electrical and computer engineering from McMaster University, Canada, in 2010 and 2013, respectively. She is currently working at VerifEye Technologies, Canada. Her research interests include image and video coding, and image processing. She was a recipient of the NSERC Alexander Graham Bell Canada Graduate Scholarship from 2011 to 2012.

PLACE
PHOTO
HERE

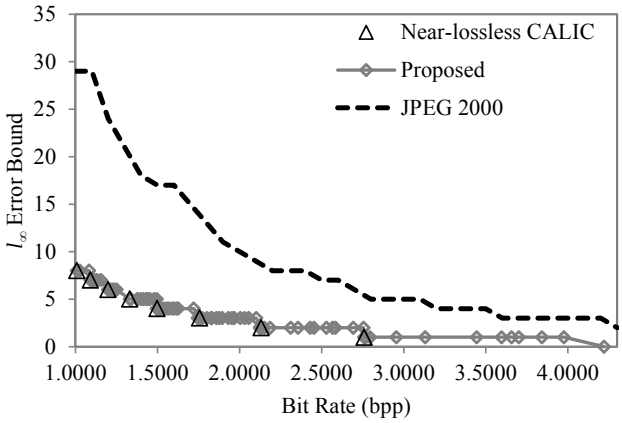
Sorina Dumitrescu received the B.Sc. and Ph.D. degrees in mathematics from the University of Bucharest, Romania, in 1990 and 1997, respectively. From 2000 to 2002 she was a Postdoctoral Fellow in the Department of Computer Science at the University of Western Ontario, London, Canada. Since 2002 she has been with the Department of Electrical and Computer Engineering at McMaster University, Hamilton, Canada, where she held Postdoctoral, Research Associate, and Assistant Professor positions, and where she is currently an Associate Professor. Her current research interests include multimedia coding and communications, network-aware data compression, multiple description codes, joint source-channel coding, signal quantization. Her earlier research interests were in formal languages and automata theory. Dr. Dumitrescu held an NSERC University Faculty Award during 2007-2012.

PLACE
PHOTO
HERE

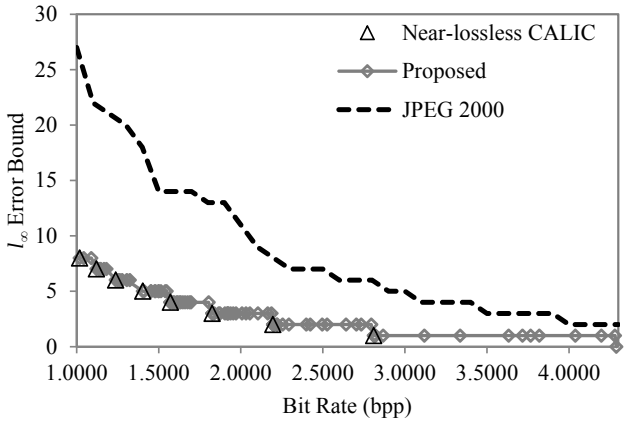
Xiaolin Wu got his B.Sc. from Wuhan University, China in 1982, and Ph.D. from University of Calgary, Canada in 1988, both in computer science. Dr. Wu started his academic career in 1988, and has since been on the faculty of University of Western Ontario, New York Polytechnic University, and currently McMaster University, where he is a professor at the Department of Electrical & Computer Engineering. His research interests include image processing, multimedia signal coding and communication, joint source-channel coding, multiple description coding, and network-aware visual communication. He has published over two hundred research papers and holds three patents in these fields. Dr. Wu is an IEEE fellow, a past associated editor of IEEE Transactions on Image Processing and on Multimedia.



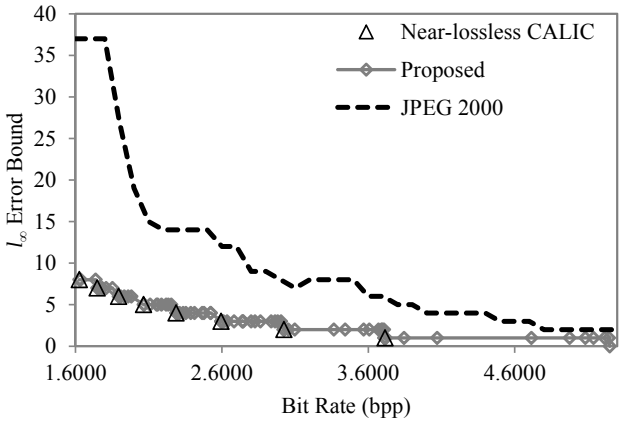
(a) Hair



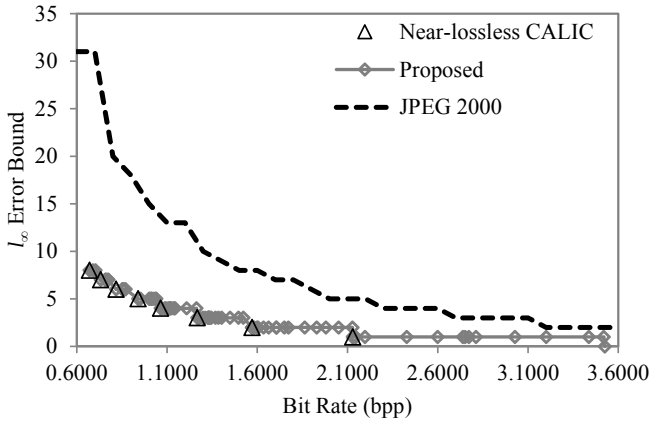
(b) Plant



(c) Flowers

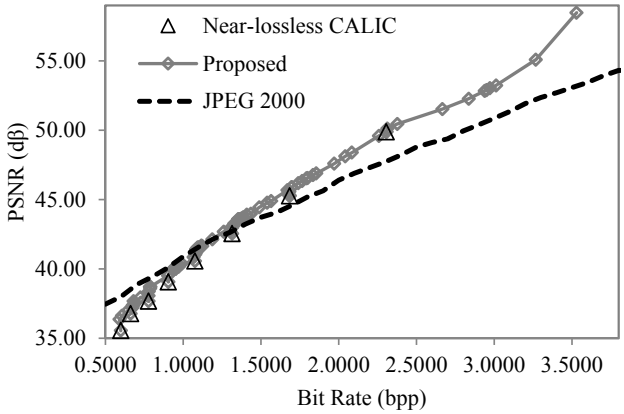


(d) Plants

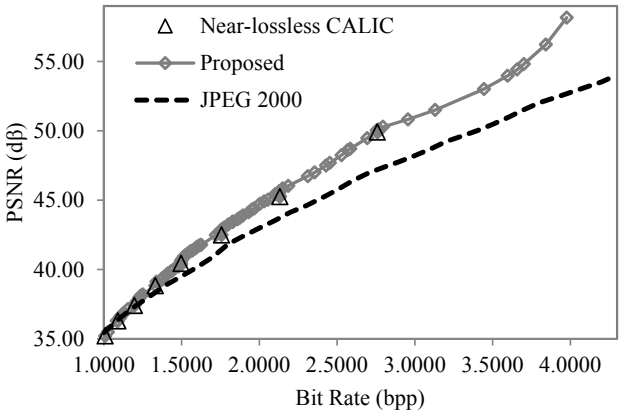


(e) Fruits

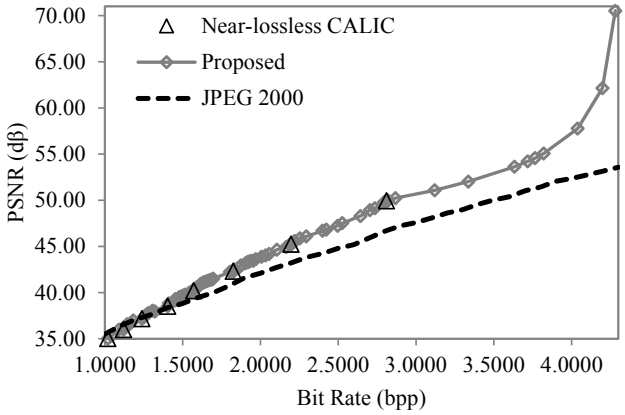
Fig. 8. ℓ_∞ error bound of test images compressed at different rates.



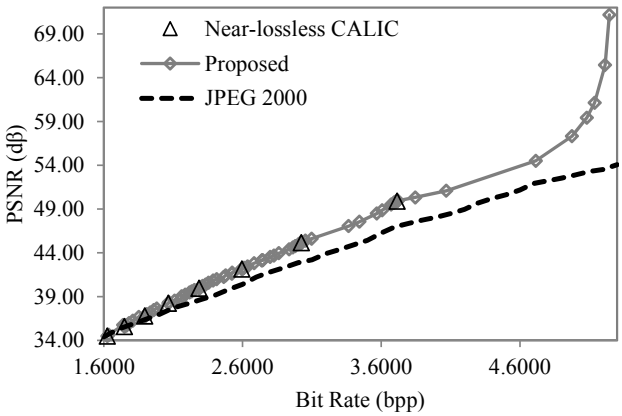
(a) Hair



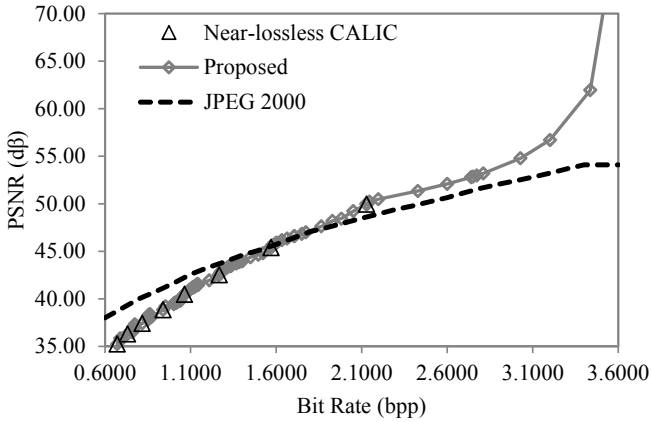
(b) Plant



(c) Flowers



(d) Plants



(e) Fruits

Fig. 9. PSNR of test images compressed at different rates.