Correspondence

Smoothness-Constrained Quantization for Wavelet Image Compression

W. Knox Carey, Sheila S. Hemami, and Peter N. Heller

Abstract—Uniform quantization of wavelet coefficients introduces perceptually disturbing artifacts by decreasing visual smoothness. In this paper, local Hölder regularity is used to quantify these changes in visual smoothness. A modified uniform quantization scheme that constrains local regularity decreases is proposed. The quantizer modification is applicable to any uniform quantization scheme and requires no additional side information to be transmitted to the decoder, unlike standard deadzone quantizers. The reconstructed images show a noticeable perceptual improvement as well as an increased PSNR relative to uniform quantization.

Index Terms—Dead zone quantization, perceptual image compression, smoothness constrained quantization, wavelet coefficient shrinkage.

I. INTRODUCTION

Algorithms based on the wavelet transform represent the current state of the art in image compression. Quantization of wavelet coefficients in lossy compression algorithms can produce perceptually disturbing artifacts in images: ringing around edges (i.e., Gibbs phenomenon) and spurious high-frequency detail in previously smooth areas. The second type of artifact occurs when there are Kronecker-delta-like features in the original image, as shown in Fig. 1, and occurs when the magnitudes of fine scale wavelet coefficients in smooth areas of the image are increased during quantization. Psychovisual studies have demonstrated that humans find compression artifacts in smooth regions more disturbing than those in edge or detailed regions [1], while errors around edges are somewhat masked [2]. As such, the focus of this work is to constrain the decreases in regularity that produce artifacts in smooth regions, rather than explicitly limiting the effect of Gibbs phenomenon.

This work analyzes changes in visual smoothness as quantified by Hölder regularity and proposes a method that mitigates perceptually disturbing decreases in local regularity. A smoothness constraint is imposed by a modified quantization scheme which reduces the probability that a coefficient will increase in magnitude. This quantizer modification requires no additional side information to be sent to the decoder, and does not rely on probabilistic models of quantization noise. Images compressed and reconstructed with this method exhibit a noticeable perceptual improvement as well as an increased PSNR relative to a uniform quantizer.

The organization of this correspondence is as follows. Section II introduces the concept of Hölder regularity and discusses how wavelet transform coefficients may be used to characterize it. Section III introduces the smoothness-constrained quantizer. Experimental results are presented in Section IV, followed by conclusions.

II. REGULARITY AND WAVELETS

This section introduces the concept of Hölder regularity to quantify changes in visual smoothness. A function $f: \mathbb{R} \rightarrow \mathbb{R}$ has Hölder regularity $\alpha = n + r$ with $n \in \mathbb{N}$ and $0 \leq r < 1$ if there exists a constant $K < \infty$ such that

$$|f^n(y) - f^n(x)| \leq K|y - x|^r \quad \forall x, y \in \mathbb{R}$$

(1)

The Hölder regularity indicates the number of continuous derivatives that a function possesses. Functions with a large Hölder exponent are smooth, while small Hölder exponents are associated with relatively rougher functions. The wavelet transform provides a practical framework in which to measure the local Hölder regularity of a function.

The dyadic wavelet transform projects a signal onto functions $\{\psi_{k,i}(x)\}$ which are derived by dilation and translation from a single mother wavelet $\psi(x)$ as $\psi_{k,i}(x) = \psi(2^k x - i)$, where the parameter $k$ denotes the scale of the function and $l$ is the translation. Let the set $\{\psi_{k,i}(x)\}$ represent dilations and translations of a compactly supported mother wavelet $\psi(x)$ with Hölder regularity $\gamma$.

The local regularity of a function $f(x)$ at a point $x_0$ may be determined from its wavelet coefficients if the mother wavelet is compactly supported [3]. Let $S$ be the set of index pairs $(k,l)$ such that for some $\epsilon > 0$, an interval $(x_0 - \epsilon, x_0 + \epsilon) \subset$ support($\psi_{k,l}$). The function has local Hölder regularity $\alpha < \gamma$ at $x_0$ if there exists a $C < \infty$ such that

$$\max_{(k,l) \in S} |w_{k,l}| \leq C 2^{-k(\gamma + 1/2)}.$$  

(2)

There is a strong correspondence between Hölder regularity and visual smoothness. If a function has high local regularity, it appears smooth. Conversely, areas with low regularity contain nonsmooth edges and textures. Quantization of wavelet coefficients can alter the local regularity of signals, introducing low regularity artifacts into previously smooth areas and smoothing abrupt changes in luminance.

III. SMOOTHNESS-CONSTRAINED QUANTIZATION

A. Effects of Quantization on Regularity

Quantization can either increase or decrease local regularity. A decrease in local regularity caused by quantization can introduce
perceptually disturbing artifacts in smooth image areas, as shown in Fig. 1. Quantization can also increase the regularity of low-regularity features such as edges, but such artifacts are less noticeable and thus more acceptable to the human visual system than decreases in local regularity. By appropriately modifying the quantization scheme, these regularity decreases can be tempered or eliminated entirely.

Let $e_{k,1}$ be the quantization error for wavelet coefficient $w_{k,1}$. To find the change in regularity associated with quantization of $w_{k,1}$, choose a number $\delta_{k,1}$ such that

$$|w_{k,1} + e_{k,1}| = C \cdot 2^{-k(\alpha+1/2)} 2^{-k/2\alpha L},$$

(3)

The Hölder regularity of the reconstructed signal after the quantization of this coefficient is $\alpha + \delta_{k,1}$. Rearranging and applying the decay inequality (2), it can be shown that if the magnitude of the coefficient is reduced by quantization, $\delta_{k,1}$ is positive, so $|\delta_{k,1}| = \delta_{k,1}$ and

$$|\delta_{k,1}| \geq \frac{1}{k} \log_2 \left( \frac{|w_{k,1}|}{|w_{k,1} + e_{k,1}|} \right)$$

(4)

On the other hand, if the magnitude of the coefficient is increased by quantization, $\delta_{k,1} < 0$ and $|\delta_{k,1}| = -\delta_{k,1}$. Therefore,

$$|\delta_{k,1}| \leq \frac{1}{k} \log_2 \left( \frac{|w_{k,1}|}{|w_{k,1} + e_{k,1}|} \right)$$

(5)

The inequalities above bound the change in regularity that may be incurred by a single coefficient. However, the local regularity of a signal in a given region $S$ is affected by the quantization of all wavelet coefficients corresponding to basis functions with support intersecting $S$. In a five scale wavelet decomposition, for example, the local regularity in every spatial location depends upon at least five different wavelet coefficients. The total change in local regularity due to quantization must take into account the quantization of all of the relevant coefficients. Let $S$ be the set of indices $(k, l)$ for which support($v_{k, l}$) $\cap$ $S$ $\neq$ $\emptyset$. The exponential on the right-hand side of (2) must decay more slowly than the magnitude of all the wavelet coefficients in $S$. The regularity in region $S$ is therefore limited by the coefficient $w_{k,1}$ which leads to the $\delta_{k,1}$ closest to $-\infty$. The total local change in Hölder regularity is therefore defined to be

$$\delta \triangleq \min_{(k, l) \in S} \delta_{k,1}.$$  

(6)

If $\delta < 0$, the local regularity and visual smoothness of the function are reduced. Likewise, the regularity and smoothness increase when $\delta > 0$.

Inequalities (4) and (5) imply that regularity is most sensitive to quantization of small wavelet coefficients. When $|w_{k,1}|$ is small, the magnitude of the quantization error $e_{k,1}$ may be comparable to the magnitude of the coefficient itself, leading to a relatively large negative $\delta_{k,1}$. If the magnitude of the coefficient $w_{k,1}$ is large, on the other hand, the relative magnitude of the quantization error will be much smaller, thus introducing a very small change in local regularity.

### B. Modifying the Uniform Quantizer to Constrain Smoothness

Consider a midtread uniform quantizer, in which the bin width is given by $\Delta$, the decision thresholds are given by $|2n+1|\Delta/2$, and the reconstruction levels are given by $n\Delta$. These quantizers have simple quantization rules and are the most commonly used quantizers, for computational simplicity. If the wavelet coefficients are modeled by Laplacian or generalized Gaussian probability density functions, the bin centroids are closer to zero than the reconstruction levels for all bins except the zero bin. Since all coefficients whose magnitudes are smaller than the bin midpoint are quantized to the midpoint, the probability of a coefficient magnitude increasing is greater than 1/2.

To counteract the effect of increasing coefficient magnitudes in a uniform quantizer, the wavelet coefficients are first passed through a nonlinear soft thresholding function $\theta$:

$$\theta(w_{k,1}) = \begin{cases} w_{k,1} - \tau \operatorname{sgn}(w_{k,1}) |w_{k,1}| \geq \tau, \\ 0 \quad |w_{k,1}| < \tau. \end{cases}$$

The threshold $\tau$ is defined to be $(\beta - 1)\Delta/2$, where $\Delta$ is the quantizer bin width and $\beta \geq 1$. The thresholded coefficients $\theta(w_{k,1})$ are then quantized with the midtread uniform quantizer. The decoder therefore requires no modification or side information. Thresholding shrinks the magnitude of any coefficients with magnitude greater than $\tau$ and replaces all others with zero. Soft thresholding followed by quantization is equivalent to using a modified uniform quantizer with zero bin width $\beta\Delta$. This smoothness-constrained quantizer is shown in Fig. 2, in which the reconstruction levels are given by $n\Delta$.

The choice of $\beta$ modifies the decision thresholds of the smoothness-constrained quantizer relative to those of the uniform quantizer, while the reconstruction levels remain the same. Note that if a Laplacian distribution is assumed for the source, $\beta$ can be chosen for the smoothness-constrained quantizer to force the reconstruction levels $n\Delta$ to be at exactly the centroid of the bins. However, the choice of $\beta$ is also affected by the changes in regularity.

The maximum amount by which a coefficient may increase in a uniform quantizer is $\Delta/2$. With a central bin expansion factor $\beta = 2$, the coefficient magnitude is reduced by $\tau_k = \Delta_k/2$ before quantization, which prevents the magnitude of the coefficient from increasing. Therefore with $\beta = 2$, the reconstructed signal is guaranteed to be at least as smooth as the original signal. In practice, choosing $\beta = 2$ tends to oversmooth edges and blur high-frequency detail. Quantizers with $\beta < 2$ prevent oversmoothing by constraining the amount by which coefficient magnitudes may increase, but cannot be guaranteed to preserve the regularity of the original signal. The value of $\beta$ must be chosen to yield an acceptable balance between smoothing and suppression of spurious low regularity artifacts.

The smoothness-constrained quantizer in Fig. 2 bears similarity to a uniform quantizer with an expanded center bin, known as a deadzone quantizer [5]. The deadzone quantizer, however, has reconstruction levels at the centers of the bins, given by $n\Delta + (\beta - 1)\Delta/2$ and the choice of $\beta$ affects both the decision thresholds and the reconstruction levels. Furthermore, it requires $\beta$ to be transmitted to the decoder. With a Laplacian source, the deadzone quantizer has a constant error between the bin centroids and reconstruction levels. 

As such, with appropriate selection of $\beta$ the smoothness-constrained quantizer can exhibit an improved performance over that of the deadzone quantizer.
Fig. 3. Reconstruction of a step edge following analysis and quantization with both a smoothness-constrained and a deadzone quantizer. The deadzone quantizer produces more overshoot and undershoot.

Fig. 4. Smoothness decreases ($\delta$) versus $\beta$ for both smoothness-constrained and deadzone quantization, measured for Lena at 16:1 and 64:1 compression.

The wavelet transform combined with soft thresholding has also been proposed for signal denoising applications [6]. In [7], this method was used to reduce quantization noise at the decoder. This approach, however, models the quantization error as additive Gaussian noise rather than using exact knowledge of the quantization errors and local regularity.

IV. EXPERIMENTAL RESULTS AND COMPARISONS

The smoothness-constrained quantizer is evaluated by comparing its performance with those of a uniform quantizer ($\beta = 1$) and a deadzone quantizer. Results from a six-level analysis, quantization, and synthesis of a one-dimensional (1-D) step edge for constant $\beta = 2$ are shown in Fig. 3, and demonstrate that the smoothness-constrained quantizer produces a sharper edge as well as less overshoot and undershoot than the deadzone quantizer. Lower values of $\beta$ produce similar though less dramatic results. However, to produce maximum PSNR, the two quantizers must use different values of $\beta$. When the optimal $\beta$ values are used, the visual smoothness as quantified by (6) is nearly equal. This is demonstrated in the simulations on images. The quantizers are evaluated on images from the USC database compressed using a 7/9 six-level biorthogonal wavelet decomposition, quantization (both smoothness-constrained and deadzone with varying $\beta$‘s) and a zero-run coder followed by a Huffman coder.

The changes in smoothness as quantified by (6) are compared for both smoothness-constrained and deadzone quantizers. For reasonable distributions such that the first quantization bins on either side of the zero bin are populated (as is the case with images), the total change in Hölder regularity can be expressed as $\delta_{sc} = (1/k) \log_2 2/\beta$ for the smoothness-constrained quantizer and $\delta_{dz} = (1/k) \log_2 (\beta + 1)/\beta$ for the deadzone quantizer, so $\delta_{dz} \leq \delta_{sc}$ and the smoothness-constrained quantizer exhibits a lower decrease in smoothness. The measured results for the Lena image at 16:1 and 64:1 compression are shown in Fig. 4, and match this prediction well. The lower decrease in smoothness for 64:1 compression results from the fact that all coefficients in the finest subbands are quantized to zero at this high compression ratio, and the largest changes in smoothness occur in scale $k = 2$ as opposed to scale $k = 1$ for 16:1 compression.

Use of a smoothness constrained quantizer in the compression algorithm produces an increase in peak signal-to-noise ratio (PSNR) relative to both a uniform quantizer and deadzone quantizer, as shown in Fig. 5. Similar results hold for higher compression ratios. The improvement relative to the uniform quantizer is due to the longer

Fig. 5. PSNR versus $\beta$ for both smoothness-constrained quantization and deadzone quantization for 16:1 compression of: (a) Lena, (b) couple, and (c) peppers.

Fig. 6. Average zero-run lengths versus $\beta$ for any quantizer with an expanded zero bin.
TABLE I
Smoothness-Constrained PSNR Improvements Relative to a Uniform Quantizer. Dead-Zone improvements are less than 0.1 dB below these values though at higher values of $\beta$

<table>
<thead>
<tr>
<th>Image</th>
<th>16:1</th>
<th>32:1</th>
<th>64:1</th>
<th>128:1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\Delta$PSNR</td>
<td>$\beta$</td>
<td>$\Delta$PSNR</td>
<td>$\beta$</td>
</tr>
<tr>
<td>lena</td>
<td>0.256</td>
<td>1.25</td>
<td>0.198</td>
<td>1.25</td>
</tr>
<tr>
<td>peppers</td>
<td>0.143</td>
<td>1.2</td>
<td>0.191</td>
<td>1.2</td>
</tr>
<tr>
<td>couple</td>
<td>0.380</td>
<td>1.25</td>
<td>0.204</td>
<td>1.25</td>
</tr>
<tr>
<td>mandrill</td>
<td>0.395</td>
<td>1.3</td>
<td>0.368</td>
<td>1.35</td>
</tr>
</tbody>
</table>

Fig. 7. Lena compressed at 64:1 using various quantizers: (a) original, (b) uniform quantizer ($\beta = 1$), PSNR = 30.2 dB, (c) smoothness-constrained quantizer ($\beta = 1.25$, PSNR = 30.4 dB, and (d) deadzone quantizer ($\beta = 1.6$), PSNR = 30.3 dB. Both (c) and (d) exhibit similar visual smoothness, which is greater than (b).
zero-runs resulting from a wider zero bin, allowing more bits to be allocated to nonzero coefficients, and is also exhibited by the deadzone quantizer. Fig. 6 shows example zero-run length versus $\beta$ for Lena at two compression ratios, and applies to any quantizer with an expanded zero bin. The improvement relative to the deadzone quantizer is due to the ability of the smoothness-constrained quantizer to place the reconstruction levels closer to the centroids of the bins. The maximum PSNR improvement of the smoothness-constrained quantizer relative to the uniform quantizer averages 0.25 dB for most images in the USC image database, with $\beta$ between 1.2 and 1.4. Several results for compression ratios ranging from 16:1 to 128:1 are shown in Table I. Deadzone quantizers also exhibit an increase in PSNR relative to a uniform quantizer, though the improvement is slightly less than that of the smoothness-constrained quantizers (less than 0.1 dB) and the maximum PSNR’s occur at optimal $\beta$’s that are higher, between 1.4 and 1.9.

Though the optimal $\beta$ to produce the maximum PSNR differs for the smoothness-constrained and deadzone quantizers, the change in smoothness as computed by (6) and plotted in Fig. 4 is similar at the optimal $\beta$’s. As such, the visual results for the smoothness-constrained and deadzone quantizers are similar when optimal $\beta$’s are used, and are better than those for the uniform quantizer, as shown in Fig. 7. However, the smoothness-constrained quantizer can produce a slight PSNR improvement and does not require $\beta$ to be known at the decoder.

V. CONCLUSIONS

Changes in visual smoothness caused by wavelet coefficient quantization are quantified mathematically by analyzing changes in the local H"{o}lder regularity. A smoothness-constrained quantizer limits the amount by which the regularity can decrease. Choosing the central quantization bin to be 1.2 to 1.4 times the width of the regular quantization stepsize produces improved visual results as well as an average PSNR increase of 0.25 dB relative to the uniform quantizer. Though a deadzone quantizer can produce similar smoothness decreases though with a larger central bin width, the PSNR performance of the smoothness-constrained quantizer will be slightly higher. The smoothness constraint can be applied to any quantization scheme, including successive approximation quantizers and lattice vector quantizers, with a computationally inexpensive encoder side modification.

REFERENCES