

OPTIMAL REQUANTIZATION-BASED RATE ADAPTATION FOR H.264

Bo Shen

Hewlett Packard Laboratories
boshen@hpl.hp.com

ABSTRACT

Bit rate adaptation is one of the most important types of video transcoding. With H.264 becoming the predominant video codec of choice in video coding and streaming, prudent rate-adaptation techniques should be developed. In this paper, we investigate certain critical points in the spectrum of rate shaping requests. We show that the selection of quantization step sizes may not have monotonic effects on rate-distortion characteristics in the transcoding sense. In other words, distortion in the regular sense can be different from the distortion in the context of transcoding in which a requantization process is carried out. We show in a generic form that careful selections of the step size can lead to much improved performance. Experiments based on both simulation and real transcoding show the effectiveness of the proposed solution.

1. INTRODUCTION

H.264 [1] (and many other compression standards) uses uniform scalar quantizer for quantization. A uniform scalar quantizer Q can be modeled as follows. Given input x , quantization output q is produced as

$$q = Q(x) = \begin{cases} \text{sign}(x) \left\lfloor \frac{|x|}{s} + \varepsilon \right\rfloor, & \frac{|x|}{s} + \varepsilon > 0 \\ 0, & \text{otherwise} \end{cases}, \quad (1)$$

where s is the quantizer step size, and ε controls the size of the deadzone. The deadzone control factor ε is often within $[0, 1/2]$.

Video transcoding involves a *requantization* process. Specifically, an input video of a transcoder contains first-generation coefficients that have been quantized with a first quantizer (generally decided by step size s_1 , and deadzone factor ε_1), the transcoder reconstructs these first-generation coefficients, then *requantizes* them by a second quantizer (s_2 , ε_2). Consider the transcoder always starts with an already-compressed input video with already quantized-coefficients, we can generalize the requant model as in Figure 1.

For the requantization process, we focus on the more common cases when $s_2 > s_1$. That is, a finer quantizing is carried out followed by a coarser quantizing. This case applies to the typical video application of high-to-low bit rate adaptation.

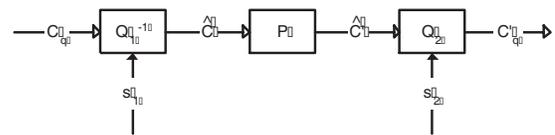


Fig. 1. Model of requantization

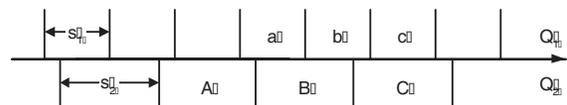


Fig. 2. Bins in requantization.

Clearly, the quantization error through the requantization process is mostly decided by the coarser quantizer. But, the quantization error can be different had the first quantization not happened (*direct quantization*), even though it uses a finer step size. Therefore, we want to evaluate the *requantization error* which is defined as the difference between the results from the direct quantization (by s_2) and the requantization (first by s_1 then by s_2). This study has significant impact on how the transcoder would select s_2 .

In [2], we have evaluated a similar problem in a specific case, that is, when $\varepsilon_1 = \varepsilon_2 = 1/2$ which applies to JPEG-to-JPEG transcoding cases. In this paper, we derive a generic framework in the next section, evaluating the unique non-monotonic property of the requantization process in general. The investigation yields optimal step-size-selection schemes, based on which, we carry out some experiments through both simulations and H.264-to-H.264 transcoding. The results are presented in Section 3. And we conclude in Section 4.

2. NON-MONOTONIC REQUANTIZATION

Assuming Q_1 and Q_2 are both uniform scalar quantizers, Figure 2 illustrates two different sets of bins generated with two different quantization steps. The original quantization step size (for Q_1) and the requantization step size (for Q_2) are s_1 and s_2 respectively.

We focus the analysis on one important aspect of the re-

quantization model, that is, *responsiveness*. Considering a set of raw coefficients C , they are *responsive* if they are first quantized into a s_1 bin which crosses the boundary of certain s_2 bins (e.g., bins a and b cross the boundaries of bins A and B, and bins B and C, respectively). Conversely, the coefficients are *unresponsive* if they are first quantized into a s_1 bin which is fully within the boundary of a s_2 bin (e.g., bin c is fully within bin C).

Since the quantization is always symmetric to zero, we focus on the investigation of non-negative coefficients for the rest of the paper without losing generality. For the sake of discussion, we denote the sets of integers, non-negative integers, positive integers as \mathbf{Z} , \mathbf{Z}^* and \mathbf{Z}^+ , respectively. Based on the quantization model, the upper and lower bounds of the interval into which a coefficient $q \in \mathbf{Z}^*$ falls are defined as follows.

$$L_q^s = (q - \varepsilon)s, U_q^s = (q + 1 - \varepsilon)s \quad (2)$$

Let $i \in C_q$, and $j \in C'_q$, The set of unresponsive coefficients is defined as:

$$\bar{C} = \{i | \forall i, j \in \mathbf{Z}, L_i^{s_1} \geq L_j^{s_2} \text{ and } U_i^{s_1} \leq U_j^{s_2}\}. \quad (3)$$

We then arrive at the following theorem.

Theorem: For uniform scalar quantizer, denote $r = s_2/s_1$, where s_1 and s_2 are the step sizes of the first quantizer and the requantizer, respectively. The magnitude of unresponsive coefficients, $i \in \mathbf{Z}$, shall fall into the following intervals:

$$\begin{aligned} \bar{I}_j &= [\bar{L}_j, \bar{U}_j] \\ &= \left[(j - \varepsilon_2)r + \varepsilon_1, (j + 1 - \varepsilon_2)r + \varepsilon_1 - 1 \right], \end{aligned} \quad (4)$$

where $j \in \mathbf{Z}$. The magnitude of the responsive coefficients $i' \in C_q$ shall fall into the following intervals:

$$\begin{aligned} \check{I}_j &= (\check{L}_j, \check{U}_j) \\ &= \left((j + 1 - \varepsilon_2)r + \varepsilon_1 - 1, (j + 1 - \varepsilon_2)r + \varepsilon_1 \right), \end{aligned} \quad (5)$$

where $j \in \mathbf{Z}$.

In the above theorem, ‘‘shall’’ means that not all intervals contain a responsive or unresponsive coefficient. Note that $\bar{U}_j - \bar{L}_j = r - 1$. When r is large, there can be multiple unresponsive coefficients within one interval. This also indicates that when $r < 1$, there are no unresponsive coefficients since $\bar{U}_j < \bar{L}_j$, so the interval is null. On the other hand, since $\check{U}_j - \check{L}_j = 1$ according to (5), there can be *at most* one responsive coefficient within one \check{I} interval.

It is further noted that the responsiveness is corresponding to the difference between a requantization process and a direct quantization process. Clearly, an unresponsive coefficient produces no difference between requantization and direct quantization. We next pay more attention to the evaluation of the unresponsive cases.

Given a set of coefficients, we are interested in the case that all coefficients are unresponsive. This corresponds to the case that requantization of these coefficients is the same as a direct quantization, that is, a *perfect* requantization.

The finding of this case can be carried out by focusing on the responsive intervals defined by (5). It is clear that if either \check{L}_j or \check{U}_j is an integer, then \check{I}_j contains no integer since $\check{U}_j - \check{L}_j = 1$. Therefore, we know that all coefficients are unresponsive if either \check{L}_j or \check{U}_j is an integer for all $j \in \mathbf{Z}^*$. Now, for \check{L}_j to be an integer for all j , we must have $r \in \mathbf{Z}$ and $r(1 - \varepsilon_2) + \varepsilon_1 - 1 \in \mathbf{Z}$. That is,

$$r = \frac{k + 1 - \varepsilon_1}{1 - \varepsilon_2} \quad (6)$$

where $k \in \mathbf{Z}^*$. Similarly, for \check{U}_j to be an integer for all j , we must have

$$r = \frac{k - \varepsilon_1}{1 - \varepsilon_2} \quad (7)$$

where $k \in \mathbf{Z}^+$.

Note that (6) and (7) represent essentially the same conclusion. Considering a provable fact that \check{I}_j must contain an integer for certain j if r is not an integer, we arrive at the following corollary.

Corollary 1: For requantization with uniform scalar quantizers, given a first quantizer step size s_1 and a requantizer step size s_2 , there are no responsive coefficients (i.e., all coefficients are unresponsive), if and only if:

$$r = \frac{k - \varepsilon_1}{1 - \varepsilon_2}, \quad (8)$$

and $r \in \mathbf{Z}^+$, where $k \in \mathbf{Z}^+$.

Since r needs to be an integer, we now further investigate integer solutions for (8). First, if both ε_1 and ε_2 are 0, it is obvious that (8) becomes: $r = k$. That is, any integer ratio of the quantization step sizes leads to no responsiveness in the requantization.

If only one of ε_1 and ε_2 is 0, for example, $\varepsilon_2 = 0$, there can be no integer solution for r in (8) since $0 \neq |\varepsilon_1| < 1$. On the other hand, if $\varepsilon_1 = 0$, we can denote ε_2 as y_2/x_2 without losing generality. (8) becomes $r = \frac{x_2 k}{x_2 - y_2}$. Solving the simple Diophantine equation reveals that $r = x_2 n$, where $n \in \mathbf{Z}^+$.

In cases of $\varepsilon_1 \neq \varepsilon_2$ and neither is 0, let us denote $\varepsilon_1 = y_1/x_1$, and $\varepsilon_2 = y_2/x_2$, where $0 < y_1 < x_1$ and $0 < y_2 < x_2$, and $\gcd(y_1, x_1) = \gcd(y_2, x_2) = 1$. Note that this fractional notation actually covers all of the ε cases for quantization schemes in all image and video compression standards. Based on this notation, the following corollary can be derived.

Corollary 1.1: For a first quantizer with a deadzone controlling factor ε_1 and a step size s_1 and a second quantizer with a deadzone controlling factor ε_2 and a step size s_2 , there is no requantization error if and only if the following is true:

$$r = x_2 n + \frac{y_1 x_2}{x_1} [(y_2 - x_2)^{-1} (\text{mod } x_2)], \quad (9)$$

$\varepsilon_1 \backslash \varepsilon_2$	0	1/6	1/4	1/3	1/2
0	n	$6n$	$4n$	$3n$	$2n$
1/6	-	$6n + 1$	-	-	-
1/4	-	-	$4n + 1$	-	-
1/3	-	$6n + 2$	-	$3n + 1$	-
1/2	-	$6n + 3$	$4n + 2$	-	$2n + 1$

Table 1. Example r that leads to perfect requantization

and $r \in \mathbf{Z}^+$, where $n \in \mathbf{Z}^*$. $(a)^{-1}(\text{mod } b)$ represents the inverse modulo of a modular b .

Proof is omitted due to space limitation. We now follow with more analysis on some actual requant cases. For r to be an integer, x_2/x_1 must be an integer. In other words, if x_1 does not divide x_2 , there can be no integer solution for r . On the other hand, it can be proved that $y_2 - x_2$ must have inverse modular of mod x_2 . Extended Euclidean algorithm [3] can be used to compute the modular inverse. Here we consider a special case when $y_2 = 1$. It can be proved that $(1 - x_2)^{-1}(\text{mod } x_2) = 1$. Therefore, (9) becomes:

$$r = x_2 n + \frac{y_1 x_2}{x_1}. \quad (10)$$

Based on the above, we obtain Table 1 for a number of combinations of more popularly used ε in requantization.

Note that Table 1 covers a broad range of deadzone control schemes that are typically used by JPEG, MPEG-1, -2, MPEG-4, H.263 and H.264. It provides significant guidance in the selection of requantization step size for rate-adaptation transcoding for content that is coded based on these standards.

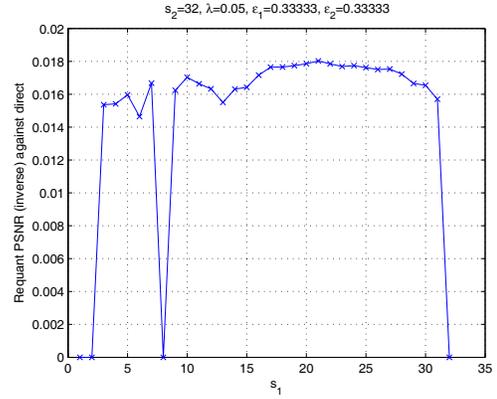
3. EXPERIMENTAL RESULT

Based on the above requantization theory, efficient rate-adapting transcoding system can be designed through optimal step size selections. Here we validate the selection scheme for some of the scenarios, first through simulations and then through real transcoding of H.264 sequences.

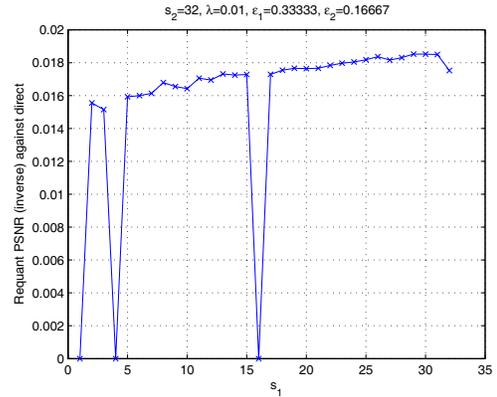
3.1. Simulation

Assuming Laplacian distribution for input coefficients to the transcoder. Several example cases of requantization with different combinations of ε_1 and ε_2 are shown in Figure 3. The shape-control factors (λ) of the distribution are also shown.

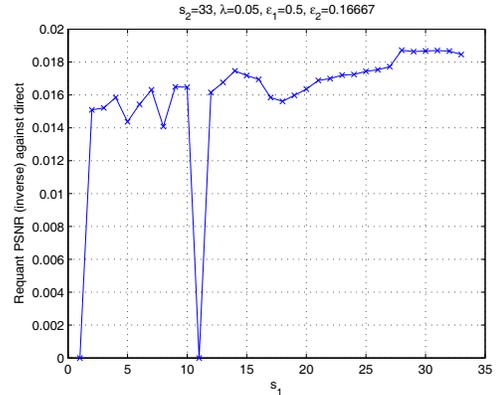
The PSNR is computed using direct quantization results as the reference. Note that since perfect requant means infinite PSNR, we plot inverse PSNR for illustration. As expected for Fig. 3(a), when $\varepsilon_1 = \varepsilon_2 = 1/3$, we observe perfect requant only at $s_1 = 2, 8, 32$, corresponding to $r = 16, 4, 1$, respectively. This is exactly what Table 1 has predicted. The same conclusion can be drawn for Figure 3(b) and (c). We have also verified the results using different λ values.



(a) H.264 intra to intra.



(b) H.264 intra to inter



(c) JPEG to H.264 inter

Fig. 3. Some example cases. Zero-valued PSNR (inverse) indicates perfect requant.

3.2. H.264 Transcoding

We now evaluate the requantization in the context of transcoding real H.264 sequences. In these experiments, we focus on the requant of intra frames in H.264 (i.e., $\varepsilon_1 = \varepsilon_2 = 1/3$). Specifically, we evaluate the quality-rate tradeoff when certain QP is picked for the transcoding. Original test sequences are coded using uniform QP1 of 16, 22, and 26, which cor-

responds to step sizes (s_1) of 4, 8, 13, respectively according to the scheme defined in H.264. Each is coded with 10 intra frames equally spaced within the total of 150 frames. To transcode them to a lower bit rate, we choose the preferred QP2 to be 28, 34 and 38 respectively. Note that due to the nonlinear mapping between QP and the actual step size in H.264 [1], the quantization step size (s_2) represented by these preferred QP2s are exactly 4 times that of s_1 , respectively. In addition, we also evaluate the performance achieved by the QPs that are in the immediate neighborhood of the preferred QP2s.

In H.264 encoding, intra frames are first spatially predicted before transformation and quantization are applied. To transcode the H.264 video, that is, requantize the transform coefficients obtained directly from a first-generation stream, we have to reuse all the spatial-prediction mode decisions. For a fair comparison, we also force the direct quantization scheme to use same mode decisions. Fig. 4 shows the results for three test sequences. PSNRs are computed using direct quantization results as the reference. The points achieved by the preferred QP2s are marked on the figures. The points to the immediate left and right of the marked QP represent the performance achieved by using QP+1 and QP-1, respectively.

These figures show, for different test scenarios, that when $s_2 = 4s_1$, the requant provides better quality with less bits than the s_2 selections in the immediate neighborhood. In general, the advantage is more significant (about 1dB better for News) for high bitrate cases, and less so (about 0.1dB better for Foreman) for low bit rate cases. Note that no *perfect* requant here is due to that direct quantization scheme is using the original frames while the transcoding scheme is using first-generation frames. Nevertheless, these experiments show that the optimal step-size selection from Table 1 indeed bring performance gain in real H.264-to-H.264 transcoding cases.

4. CONCLUSION

This paper has revealed a unique property of the requantization process: certain selections of the step size used by the requantizer lead to perfect reconstruction comparing to direct quantization. Based on concfor mathematic foundation, we have identified such step sizes for scalar quantizer with different deadzone sizes. Both simulation- and H.264-transcoding-based experiments validate this discovery. The solution can be applied to a variety of different transcoding scenarios other than just H.264-to-H.264 transcoding.

5. REFERENCES

[1] Joint Video Team of ISO/IEC MPEG and ITU-T VCEG, "Joint final committee draft of joint video specification (ITU-T Rec. H.264 ISO/IEC 14496-10 AVC)," Aug. 2002.

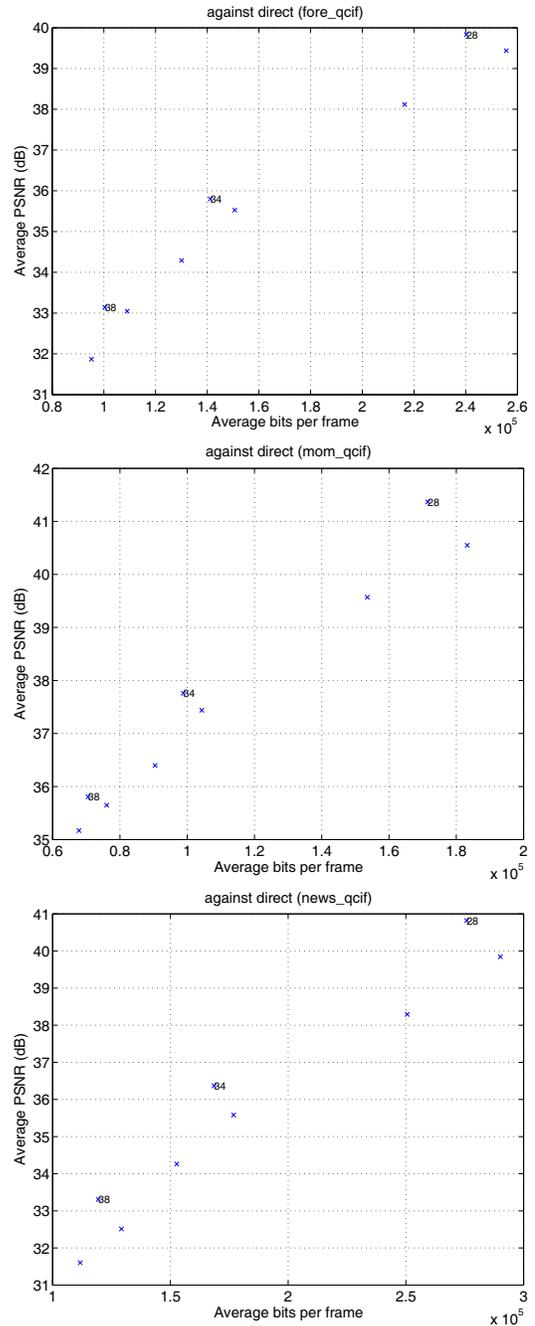


Fig. 4. Requant for H.264 intra frames

[2] B. Shen, "Modeled analysis on requantization errors," *Proc. IEEE Workshop on Multimedia Signal Processing*, Oct. 2005.

[3] K.H. Rosen, *Elementary Number Theory and its applications*, Addison Wesley Longman, Inc, Fourth Edition, 2000.