RATE-DISTORTION MODELING OF SCALABLE VIDEO CODERS

Min Dai, Dmitri Loguinov

Texas A&M University
College Station, TX 77843 USA
min@ee.tamu.edu, dmitri@cs.tamu.edu

Hayder Radha

Michigan State University
East Lansing, MI 48824 USA
radha@egr.msu.edu

ABSTRACT

After the emergence of numerous Internet streaming applications, rate-distortion (R-D) modeling of scalable video encoders has become an important issue. In this paper, we examine the performance of existing R-D models in scalable coders by using the example of MPEG-4 FGS and PFGS, and propose a novel R-D model based on approximation theory. Experimental results demonstrate that the proposed model is very accurate and significantly outperforms those in previous work.

1. INTRODUCTION

Since R-D curves are central to a wide range of video applications (e.g., optimal bits allocation [15], constant quality control [17], etc.), they continuously attract significant attention in research literature. R-D models can be classified into two categories based on the theory they apply: models based on Shannon's rate-distortion theory [5] and those derived from high-rate quantization theory [1]. The former assumes that sources are coded using very long (infinite) blocks, while the latter assumes that the encoding rate is arbitrarily high [14]. These two theories are complementary and, as shown in [6], converge to the same lower bound $D \sim e^{-\alpha R}$ when the input block size goes to infinity.

However, since block length cannot be infinite in real coding systems, it is widely recognized that classical rate-distortion theory is often not suitable for accurate modeling of actual R-D curves [14]. In addition, current transform-based encoders achieve high compression ratios and often produce low-bitrate signals [13], which means that the high-bitrate assumption of the quantization theory no longer holds [2], [13]. Therefore, adjustable parameters are often incorporated into the theoretical R-D models to keep up with the complexity of coding systems and the diversity of video sources (e.g., [9], [15]).

Recall that most current R-D models are built for images or *non-scalable* video coders. The lack of scalable R-D modeling is partially addressed in [3], which introduces a precise distortion model $D(\Delta)$ for scalable coders; however, the issue of deriving a complete R-D model remains open. To better understand R-D curves of scalable coding systems and further develop useful tools for streaming applications, in this paper, we complete the work of [3] and examine R-D models from a different perspective. We first derive a distortion model based on approximation theory and then incorporate the ρ-domain bitrate model [10] into the final result. We also show that the unifying ρ-domain model is very accurate in both Fine Granular Scalability (FGS) [12] and Progressive FGS (PFGS) [16] coders, where the former uses the enhancement layer only for reconstruction purposes while the latter also utilizes it for motion prediction.

Our work demonstrates that distortion D can be modeled by a function of both bitrate R and its logarithm $\log R$:

$$D = \sigma_x^2 - (a \log^2 R + b \log R + c)R, \tag{1}$$

where σ_x^2 is the variance of the source and $a - c$ are constants.

This paper is organized as follows. In Section 2, we briefly overview current R-D models and show their performance in scalable coders. In Section 3, we analyze distortion in scalable coders and derive a novel R-D model from the angle of approximation theory. We also demonstrate the accuracy of the proposed R-D model in various scalable video sequences. Section 4 concludes this paper.

2. RELATED WORK AND MOTIVATION

In this section, we give a brief overview of related work and analyze the applicability of current R-D models to scalable coders.

2.1. Related Work

In rate-distortion theory, there are no explicit R-D models, but only upper and lower bounds for general sources [11]:

$$Q2^{-2R} \le D(R) \le \sigma_G^2 2^{-2R}, \tag{2}$$

where Q is the entropy power and σ_G^2 is the variance of a Gaussian distributed source. In contrast, there are two kinds of lower bounds in high-rate quantization theory [8]: the minimum distortion $D_1(N)$ attainable for a constrained number of quantization levels N, and the minimum distortion $D_2(R)$ attainable for a constrained bitrate R. However, in both quantization and rate-distortion theory, D can be expressed as an exponential function of bitrate R [6]:

$$D(R) \sim Ke^{-\alpha R}, \tag{3}$$

where parameters K, $\alpha > 0$ are unspecified constants. Model (3) is rarely used in practice and many video applications often rely on its refinement [11], [15]:

$$D(R) = \gamma\varepsilon^2\sigma_x^2 2^{-2R}, \tag{4}$$

where γ is the correlation coefficient of the source and ε^2 is a source-dependent scaling parameter (1.4 for Gaussian, 1.2 for Laplacian, and 1 for uniform sources).

Based on approximation theory, Cohen *et al.* [2] derive an R-D bound for wavelet-based compression schemes. This is the first R-D bound that includes both bitrate R and $\log R$:

$$D(R) \le CR^{-2\gamma}(\log R)^{2\gamma}, \tag{5}$$

This work was supported in part by NSF grant ANI-0312461.

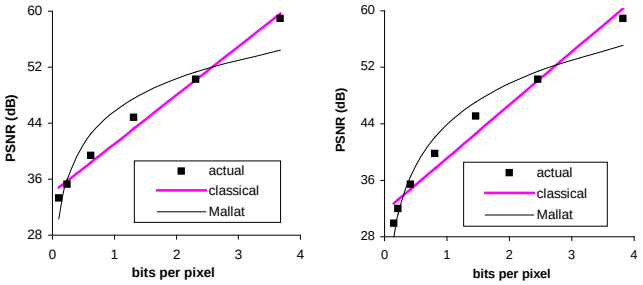

Fig. 1. The actual R-D curve and its estimations for frame 3 in FGS-coded CIF Foreman (left). The same simulation for frame 171 in the same sequence (right).

where constant C and parameter γ are both positive. Since this bound is specifically developed for wavelet-based coding schemes, Mallat *et al.* [13] extend it to transform-based low bitrate images:

$$D(R) = CR^{1-2\gamma}, \tag{6}$$

where $\gamma \approx 1, C > 0$, and the parameters are adjusted with respect to practical coding settings.

2.2. Motivation

A typical scalable coder includes one base layer and one or more enhancement layers. Notice that in scalable streaming applications, the server is often concerned with the bitrate R of the *enhancement* layer where R varies from very low bitrate (e.g., less than 0.5 bit/pixel) to high bitrate (e.g., 4 bits/pixel) depending on streaming conditions [15], [17]. Thus, we examine the accuracy of current R-D models for scalable coders, with R representing the bitrate of the enhancement layer.

Without loss of generality, we use peak signal-to-noise ratio (PSNR) to measure the quality of video sequences. With the PSNR measure, it is well-known that the classical model (4) (labeled as "classical" in all figures) becomes a linear function of coding rate R:

$$PSNR = c \log R + d, \tag{7}$$

where c and d are constants. Fig. 1 shows that PSNR is linear with respect to R only when the bitrate is sufficiently high and also that model (6) (labeled as "Mallat") has much higher convexity than the actual R-D curve. We observed similar results in other scalable sequences (not shown for brevity). While Fig. 1 is not an exhaustive demonstration, it is not surprising that current popular R-D models (4) and (6) are not directly applicable to scalable coders due to their large range of R. Given this discussion, we find that the research community would benefit from an R-D model that can precisely describe R-D curves of real scalable coders. We accomplish this task in the next section.

3. R-D MODEL FOR SCALABLE CODERS

3.1. Preliminaries

It is well known that in an ideal orthogonal transform-based coding system, the distortion in the spatial domain is the same as that in the transform domain [11]. Furthermore, recall that the distortion in an ideal transform-based video coder is mostly introduced by quantization errors [11]. Since uniform quantizers are widely applied to video coders due to their asymptotic optimality [7], we show the lower bound on distortion in quantization theory assuming seminorm-based distortion measures (e.g., mean square error (MSE)) and uniform quantizers.

If X, $\hat{X}$ are k-dimensional vectors and the distortion between X and $\hat{X}$ is $d(X, \hat{X}) = ||X - \hat{X}||^r$ (where $||\cdot||$ is a seminorm in k-dimensional Euclidean space and $r \geq 1$), the minimum distortion for uniform quantizers is [8]:

$$D = \frac{k}{k+r} \left(\frac{V_k}{\Delta} \right)^{-r/k}, \tag{8}$$

where Δ is the quantization step, $V_k = \dfrac{2\pi^{k/2}}{k\Gamma(k/2)}$, and Γ is the Gamma function. When $r = 2, k = 1$, we obtain the popular MSE formula for uniform quantizers:

$$D = \frac{\Delta^2}{\beta}, \tag{9}$$

where β is 12 if the quantization step is much smaller than the signal variance [9]. However, this assumption is not always valid in real coders and β often becomes an adjustable parameter [9]. In contrast to many previous studies based on rate-distortion and quantization theory [9], [10], in what follows, we investigate the distortion from the perspective of approximation theory and derive a novel R-D model using analytical tools not commonly available in related work.

3.2. Distortion Analysis

Assume that signal X is transformed into signal U by an orthogonal transform, which later becomes $\hat{U}$ after quantization. Since a midtread uniform quantizer is commonly used in video coders, coefficients between $(-\Delta, \Delta)$ are set to zero, where Δ is the quantization step. We call the coefficients that are larger than Δ *significant*.

As we stated earlier, distortion D between X and the reconstructed signal $\hat{X}$ equals that between U and $\hat{U}$ [11]. In the transform domain, distortion D consists of two parts: 1) distortion D_i from discarding the insignificant coefficients in $(-\Delta, \Delta)$; and 2) distortion D_s from quantizing the significant coefficients (i.e., those that have larger values than Δ). Given this notation, we have the following lemma.

Lemma 1. *Assuming that the total number of transform coefficients U is N and the number of significant coefficients is M, MSE distortion D is:*

$$D = \frac{1}{N} \sum_{|u|<\Delta} |u|^2 + \frac{M}{N} \frac{\Delta^2}{12}, \tag{10}$$

where Δ is the quantization step.

Proof. It is easy to understand that distortion D_i is directly the summation of the squares of insignificant coefficients:

$$D_i = \sum_{|u|<\Delta} |u|^2. \tag{11}$$

Since the high-resolution quantization hypothesis applies to M significant coefficients [13], their average distortion is $\Delta^2/12$ and thus their total distortion D_s is:

$$D_s = \sum_{|u|\geq\Delta} |u - \hat{u}|^2 = \frac{M\Delta^2}{12}. \tag{12}$$

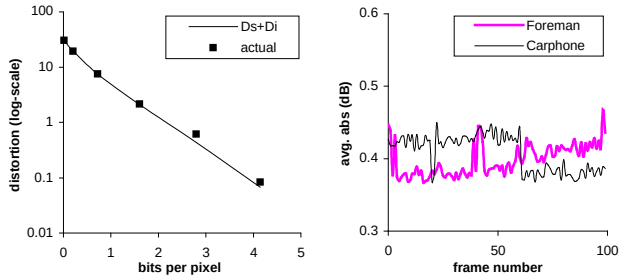

Fig. 2. Actual distortion and the estimation of model (13) for frame 3 in FGS-coded CIF Foreman (left). The average absolute error between model (10) and the actual distortion in FGS-coded CIF Foreman and CIF Carphone (right).

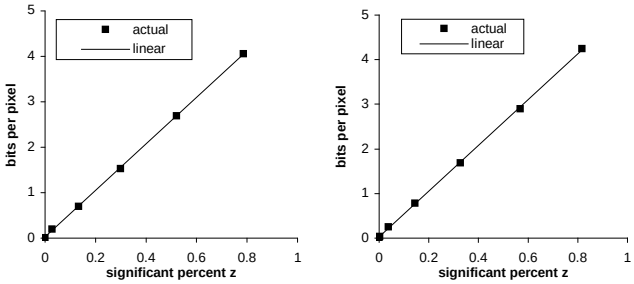

Fig. 3. Bitrate estimation of the linear model $R(z)$ for frame 0 in FGS-coded CIF Foreman (left) and frame 6 in PFGS-coded CIF Coastguard (right).

Therefore, the average distortion for each coefficient D is:

$$D = \frac{D_i + D_s}{N}, \tag{13}$$

which, combined with (11)-(12), leads to the result in (10). □

We define $z = M/N$ to be the percentage of significant coefficients in the following discussion. In Fig. 2, the left side shows an example of actual distortion D and simulation results of model (10) for frame 3 in FGS-coded CIF Foreman, and the right side shows the average absolute error between model (10) and the actual distortion in FGS-coded CIF Foreman and Carphone sequences. As we observe from the figure, model (10) is very accurate and follows the actual distortion well. Furthermore, a less-than-0.5 dB average error is rather minor since a video sequence usually has quality above 30 dB.

3.3. R-D Modeling

To improve the unsatisfactory accuracy of current R-D models in scalable coders, we derive an accurate R-D model based on source statistical properties and a recent ρ-domain model. He *et al.* [10] proposed a unified ρ-domain model to estimate the bitrate of image and non-scalable video coders, in which bitrate R is a linear function of the percentage of significant coefficients z in each video frame. We extensively examined the relationship between R and z in various video frames and found this linear model holds very well for scalable coders. Fig. 3 demonstrates two typical examples of the actual bitrate R and its linear estimation in FGS and PFGS video frames. Using the ρ-domain model [10], we have our main result as following.

Theorem 1. *The distortion of scalable video coders is given by:*

$$D = \sigma_x^2 - \left(a \log^2 R + b \log R + c\right) R, \tag{14}$$

for some constants $a - c$.

Proof. Notice that the transform coefficients U of scalable coders often follow a mixture Laplacian distribution with density [3]:

$$f(x) = p\frac{\lambda_0}{2}e^{-\lambda_0|x|} + (1-p)\frac{\lambda_1}{2}e^{-\lambda_1|x|}. \tag{15}$$

During the discussion, we first use pure Laplacian-distributed sources for simplicity (i.e., $f(x) = \frac{\lambda}{2}e^{-\lambda|x|}$), and then obtain the final version of R-D model.

Since the coefficients inside the zero bin $(-\Delta, \Delta)$ are set to zero after quantization, the average distortion of the insignificant coefficients is:

$$\begin{aligned}\frac{D_i}{N} &= \int_{-\Delta}^{\Delta} x^2 \frac{\lambda}{2}e^{-\lambda|x|}dx \\ &= \frac{2}{\lambda^2} - [(\Delta + \frac{1}{\lambda})^2 + \frac{1}{\lambda^2}]e^{-\lambda\Delta},\end{aligned} \tag{16}$$

where N is the total number of coefficients and λ is the shape parameter of the Laplacian distribution. Notice that for Laplacian distributed sources, the percentage of significant coefficients is:

$$z = 1 - 2\int_0^{\Delta} \frac{\lambda}{2}e^{-\lambda x}dx = e^{-\lambda\Delta}. \tag{17}$$

Thus, distortion D in (10) becomes:

$$D = \frac{D_i}{N} + \frac{z\Delta^2}{12} = \frac{2}{\lambda^2} - \zeta e^{-\lambda\Delta} + \frac{e^{-\lambda\Delta}\Delta^2}{12}, \tag{18}$$

where $\zeta = (\Delta + \frac{1}{\lambda})^2 + \frac{1}{\lambda^2}$.

Next, recall that bitrate $R(z) = \gamma z$, where γ is some source-dependent constant [10]. Thus, we express Δ in terms of rate R:

$$\Delta = -\frac{1}{\lambda}\log\frac{R}{\gamma}, \quad 0 < \frac{R}{\gamma} \le e^{-\lambda}. \tag{19}$$

Therefore, combining (19) with (18), distortion D is a function of R and $\log R$:

$$D = \frac{2}{\lambda^2} - \frac{11\tau^2 + 24\tau + 24}{12\lambda^2}\frac{R}{\gamma}, \tag{20}$$

where $\tau = \log\gamma - \log R$. We also notice that:

$$D = \begin{cases} \frac{2}{\lambda^2} = \sigma_x^2, & R = 0 \\ 0, & R \ge e^{-\lambda}\gamma \end{cases} \tag{21}$$

where σ_x^2 is the variance of the source. This observation makes perfect sense since distortion D should not be larger than σ_x^2 [5] and should equal zero when $R = e^{-\lambda}\gamma$ (i.e., the quantization step $\Delta = 1$ and there is no loss of information).

An R-D model for a scalable coder is simply a linear combination of (20), with corresponding probability p and distribution parameters λ_0, λ_1 as shown in (15). After absorbing the various constants, we have the desired result in (14). □

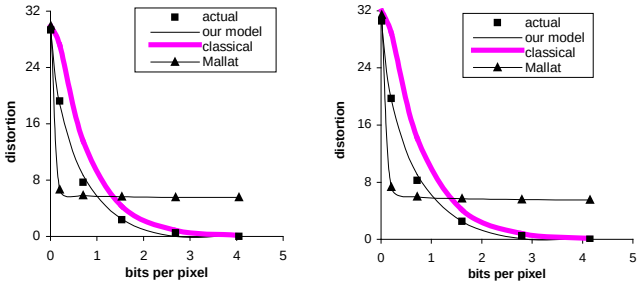

Fig. 4. Actual R-D curves and their estimations for frame 0 (left) and frame 3 in FGS-coded CIF Foreman (right).

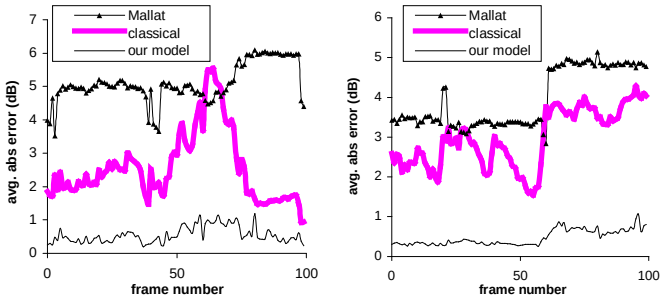

Fig. 5. The average absolute error in FGS-coded CIF Foreman (left) and CIF Carphone (right).

3.4. Experimental Results

We apply the proposed model (14) to various scalable video frames to evaluate its accuracy. Fig. 4 shows two examples of R-D curves for I (left) and P (right) frames of FGS-coded CIF Foreman. As shown in the figure, the low bitrate model (6) tends to underestimate distortion in general and saturates when bitrate R is large. Fig. 4 also shows that while the classical model (4) over-estimates the actual R-D curves, our model (14) tracks them with very high precision.

To better understand the estimation accuracy of the proposed model (14), we further compare it to models (4) and (6) in four scalable video sequences. All results shown in this paper utilize videos in the CIF format with the base layer coded at 128 kb/s and 10 frames/s. We contrast the performance of the proposed model with that of the other two models in FGS-coded Foreman and Carphone in Fig. 5. Additionally, Fig. 6 shows the same comparison in PFGS-coded Coastguard and Mobile. As both figures show, model (14) keeps the average absolute error quite low compared to that of the other models. Additional experimental results (not shown here due to a lack of space) demonstrate that (14) significantly outperforms other operational R-D models in a wide variety of scalable sequences.

4. CONCLUSION

This paper analyzed the distortion of scalable coders and proposed a novel R-D model from the perspective of approximation theory. Given the lack of R-D modeling of scalable coders, we believe this work will benefit both Internet streaming applications and theoretical discussion in this area.

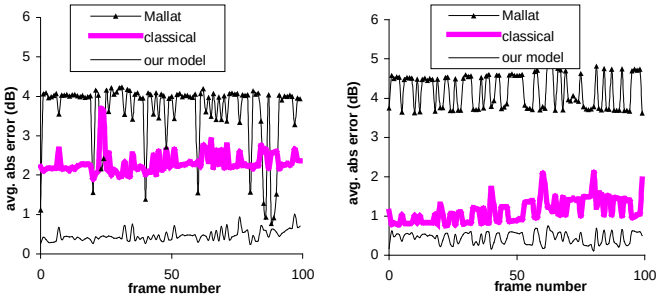

Fig. 6. The average absolute error in PFGS-coded CIF Coastguard (left) and CIF Mobile (right).

5. REFERENCES

[1] W. R. Bennett, "Spectra of Quantized Signals," *Bell sys. Tech. Journal*, vol. 27, 1948.

[2] A. Cohen, I. Daubechies, O. G. Guleryuz, and M.T. Orchard, "On the Importance of Combining Wavelet-based Nonlinear Approximation with Coding Strategies," *IEEE Trans. on Information Theory*, vol. 48, July 2002.

[3] M. Dai, D. Loguinov, and H. Radha, "Statistical Analysis and Distortion Modeling of MPEG-4 FGS," *IEEE ICIP*, Sept. 2003.

[4] R. A. Devore, B. Jawerth, and B. J. Lucier, "Image Compression through Wavelet Transform Coding," *IEEE Trans. Information Theory*, vol. 38, Mar. 1992.

[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & SONS, INC. New York, 1991.

[6] A. Gersho, "Asymptotically Optimal Block Quantization," *IEEE Trans. on Information Theory*, vol. IT-23, 1979.

[7] H. Gish and J. N. Pierce, "Asymptotically Efficient Quantizing," *IEEE Trans. on Information Theory*, vol. IT-14, Sept. 1968

[8] R. M. Gray, *Source coding theory*, Boston: Kluwer Academic Publishers, 1990.

[9] H.-M. Hang and J.-J. Chen, "Source Model for Transform Video Coder and Its Application —Part I: Fundamental Theory," *IEEE Trans. on CSVT*, vol. 7, April 1997.

[10] Z. He and S. K. Mitra, "A Unified Rate-Distortion Analysis Framework for Transform Coding," *IEEE Trans. on CSVT*, vol. 11, Dec. 2001.

[11] N. Jayant and P. Noll, *Digital Coding of Waveforms*, Englewood Cliffs, NJ: Prentice Hall, 1984.

[12] W. Li, "Overview of Fine Granularity Scalability in MPEG-4 Video Standard," *IEEE Trans. on CSVT*, vol. 11, No. 3, March 2001.

[13] S. Mallat and F. Falzon, "Analysis of Low Bit Rate Image Transform Coding," *IEEE Trans. on Signal Processing*, vol.46, April 1998.

[14] A. Ortega and K. Ramchandran, "Rate-Distortion Methods for Image and Video Compression," *IEEE Signal Processing Magazine*, Nov. 1998.

[15] Q. Wang, Z. Xiong, F. Wu, and S. Li, "Optimal Rate Allocation for Progressive Fine Granularity Scalable Video Coding," *IEEE Signal Processing Letter*, vol.9, Feb. 2002.

[16] F. Wu, S. Li, and Y.-Q. Zhang, "A Framework for Efficient Progressive Fine Granularity Scalable Video Coding," *IEEE Trans. on CSVT*, vol. 11, No. 3, Mar. 2001.

[17] X. J. Zhao, Y. W. He, S. Q. Yang, and Y. Z. Zhong, "Rate Allocation of Equal Image Quality for MPEG-4 FGS Video Streaming," *Packet Video'02*, April, 2002.