ANALYSIS AND DISTORTION MODELING OF MPEG-4 FGS

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

In this paper, we analyze statistical and rate-distortion (R-D) properties of MPEG-4 *Fine-Granular Scalability* (FGS), which has recently become an important scalable compression framework and a de-facto standard for Internet video streaming. We first propose a novel statistical model of DCT residue that accurately captures the properties of the input to the MPEG-4 FGS enhancement layer. Our results show that FGS residue concentrates a lot of probability mass near zero and cannot be accurately modeled by Gaussian or Laplacian distributions. We then model the distortion of each bitplane based on the proposed statistical framework and further demonstrate that our R-D model significantly outperforms current distortion models.

1. INTRODUCTION

Internet video streaming is an important research area in networking and video communities. To provide an error-resilient and bandwidth scalable solution to Internet video applications, *Fine Granular Scalability* (FGS) has recently been chosen as the streaming profile of the ISO/IEC MPEG-4 standard [11], [13]. Due to the inherent nature of rate control in the base layer, multi-layered encoders often produce base layers with highly fluctuating visual quality [15], [16]. In order to reduce quality fluctuation and match the video sending rate to the capacity of the network channel during streaming, the server often must rely on accurate estimation of the *rate-distortion* (R-D) curve of the video to decide how to scale the enhancement layer.

Recall that the FGS layer contains the DCT residue left over from the base layer, which means that the distortion in the FGS layer *alone* describes the distortion of the combined signal at the receiver. Therefore, for FGS-coded sequences, R-D modeling of the enhancement layer is sufficient to describe the visual quality observed by the user and has been repeatedly used in the past to achieve constant-quality scaling during transmission [14], [15], [16].

There are two approaches for estimating the R-D curve of a video encoder: the empirical approach and the analytical approach. The *empirical* approach is to construct the R-D curve based on interpolating between several sampled values of rate and distortion [7], [15], [16]. The *analytical* approach is to build a mathematical model of the source and/or encoder by analyzing statistical properties of the video data [4], [5]. Although the empirical approach is generally easy to apply, it does not give us much insight into the video coding process and its high computational requirements during streaming typically place a burden on streaming servers.

On the other hand, current closed-form analytical approaches develop closed-form solutions only for certain types of distributions (e.g., memoryless Gaussian) [6], and thus are not very accurate on most real input sequences [14]. Even though additional (heuristic) parameters estimated from the actual data can be added to obtain more accurate R-D curves [4], [5], no currently available closed-form model can capture all of the complexities of a real encoder. Furthermore, present analytical approaches are mostly developed for non-scalable video and are applied at the *base* layer [4], [5]; no specific work has been done on R-D modeling of FGS for Internet streaming applications.

There are many applications of R-D modeling of FGS (including R-D optimizations during streaming and constant-quality rate adaptation), which we consider to be beyond the scope of this paper. Our primary goal in this paper is to *understand* statistical properties of DCT residue and *study* the bitplane-coding process of the FGS enhancement layer. Our secondary goal is to derive an accurate distortion model for each bitplane since we find that the amount of work done in this important direction still remains rather scarce.

In this work, we study the properties of MPEG-4 FGS [11], [13] and propose a novel model that describes the statistical features of the input to the enhancement layer of MPEG-4 FGS. Based on this analysis, we subsequently build a distortion model for bitplane coding, which is significantly more accurate than the existing methods for a variety of FGS video sequences.

This paper is organized as follows. In section 2, we analyze and model statistical properties of the input to MPEG-4 FGS. Section 3 provides the analysis of bitplane coding and describes the proposed distortion model. Section 4 concludes this paper.

2. STATISTICAL MODEL OF DCT RESIDUE

For successful R-D modeling, correct estimation of statistical properties of source data is certainly an important factor. The enhancement layer input to the FGS encoder is the DCT residue between the original image and the reconstructed image in the base layer. Thus, we start with modeling the DCT residue and address the distortion issue in the following sections.

2.1. Statistical Properties

Gaussian and Laplacian (double exponential) distributions are the two most popular statistical models for DCT coefficients ([1], [6]) and DCT residue (e.g., [14]). However, it is possible that these models are widely applied only because of their mathematically tractability rather than their accuracy in modeling the actual data. We investigate this issue below and find that the Laplacian model

To appear in *Proc. IEEE ICIP, September 14-17, 2003, Barcelona, Spain*

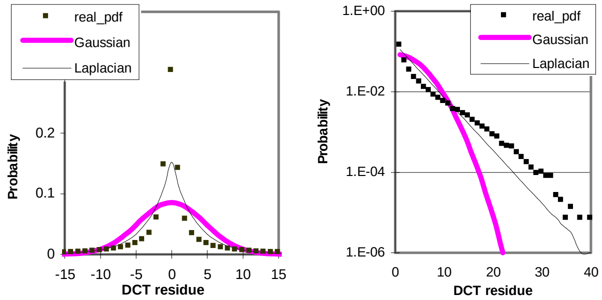

Fig. 1. The PMF of DCT residue with Gaussian and Laplacian estimations (left). Logarithmic scale of PMFs for the positive residue (right). Frame 0 of Foreman CIF coded at 10 fps and 128 kb/s in the base layer.

tracks the real DCT residue much better than the Gaussian model; however, since FGS data contains a pronounced peak at zero, a single Laplacian distribution is insufficient to completely describe statistical characteristics of DCT coefficients in the FGS layer.

To analyze the statistical properties of real DCT residue, we examined different FGS coded sequences. An example of what we observed is shown in Fig. 1. As Fig. 1 (left) shows, both the Gaussian and Laplacian distributions fail to accurately model the sharp peak in the center of the *probability mass function* (PMF) of the DCT residue. Fig. 1 (right) shows the logarithmic scale of the real PMF for the positive residue together with that of the Gaussian and Laplacian estimations. From this figure, we can see that the tail of the Gaussian distribution decays too quickly and the Laplacian distribution cannot describe the "bending" shape of the real PMF.

Also notice that in Fig. 1 (right), the tail of the log-scaled PMF of DCT residue is approximately a straight line, which means that the *tail* of the histogram can be modeled by an exponential distribution (recall that straight lines on log scale are exponential functions); however, the central part of the PMF (the peak) *cannot* be modeled by the *same* exponential distribution. To capture the sharp peaks and heavy tails, we next propose a *mixture* Laplacian model, which is a linear combination of two Laplacian distributions.

2.2. Mixture Laplacian Model

Consider the value of DCT residue as a random variable X and define a hidden state S that decides from which of the two Laplacian distributions X is drawn. Label the two states with binary numbers 0 (small variance) and 1 (large variance). Hence, the PMF of X is a mixture of two Laplacian distributions:

$$\begin{aligned} p(x) &= \sum_{n=0,1} p(S=n)p(x|S=n) \\ &= \sum_{n=0,1} p(S=n)\frac{\lambda_n}{2}e^{-\lambda_n|x|} \\ &= p\frac{\lambda_0}{2}e^{-\lambda_0|x|} + (1-p)\frac{\lambda_1}{2}e^{-\lambda_1|x|}, \end{aligned} \tag{1}$$

where λ_0 and λ_1 are the shape parameters of the two Laplacian distributions. The small-variance conditional PMF $p(x|S{=}0)$ concentrates the mass near zero, whereas the large-variance conditional PMF $p(x|S{=}1)$ spreads out the rest of the mass across larger values. It should be pointed out that our model is more than a simple curve fitting method; we use the Expectation-Maximization (EM) algorithm to obtain the Maximum-likelihood (ML) estimate for parameters $\{\lambda_1, \lambda_2, p(S{=}0)\}$.

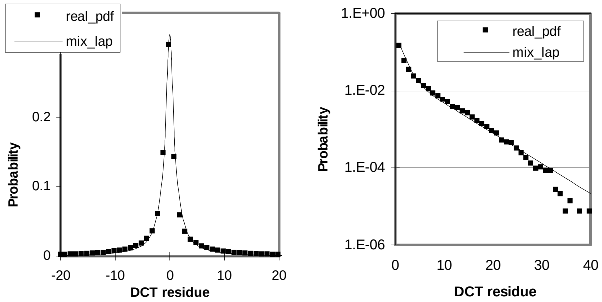

Fig. 2. Real PMF and mixture Laplacian (left). Tails on logarithmic scale of mixture Laplacian and real PMF (right). Frame 0 of Foreman CIF coded at 10 fps and 128 kb/s in the base layer.

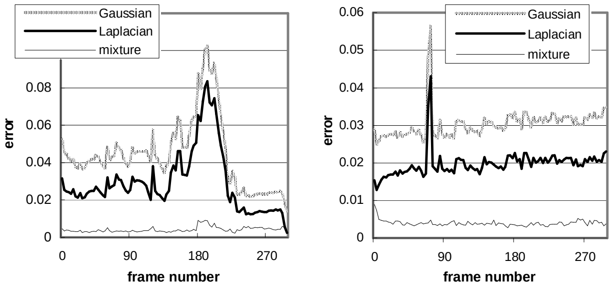

Fig. 3. The weighted absolute error of Gaussian estimation, Laplacian estimation and the mixture Laplacian estimation in Foreman CIF (left) and Coastguard CIF (right).

Fig. 2 (left) shows the mixture-Laplacian estimation and the real PMF of the DCT residue for frame 0 in Foreman CIF coded at 10 fps and 128 kb/s in the base layer. Fig. 2 (right) is the logarithmic scale of the mixture Laplacian and the real PMF. As illustrated in the figures, the mixture Laplacian distribution fits very well both the peak and the tail. The discrepancy at the end of the tail typically does not affect the accuracy of source modeling since very few samples are contained there (only 0.04% in frame 0).

We also illustrate the weighted absolute error (i.e., the absolute error times the real PMF at each value of DCT residue) of these models for Foreman CIF and Coastguard CIF in Fig. 3. Both sequences are coded at 10 fps and 128 kb/s in the base layer.

Experimental results show that straightforward application of classical (e.g., Gaussian and Laplacian) statistical models to DCT residue in FGS does not necessarily lead to accurate estimation. However, the mixture-Laplacian distribution follows over 99% of the real data with exceptional accuracy.

3. DISTORTION MODEL FOR BITPLANE CODING

There are two major difficulties in modeling the distortion using traditional rate-distortion theory. First, many sources possess such complicated statistical properties that there are no closed-form models for them, and sometimes sources even exhibit nonstationarity. Second, traditional rate-distortion theory often relies on certain approximations to build mathematically tractable R-D functions, which in turn do not model real R-D curves well. For example, the $i.i.d.$ (independent, identically distributed) source assumption discards the correlation structure existing in real coders, while the high-resolution assumption (i.e., the histogram of the source data is constant in each quantization bin [10]) does not hold for large quantization steps Δ. Hence, any direct application of classical R-D models is often not accurate and requires estimations of several empirical parameters as mentioned in the introduction.

3.1. Previous Distortion Models

In the traditional rate-distortion theory, the distortion function is established based on the mutual information as a measure of the transmission of information from the source to the user [6]. Straightforward derivations using the classical model [4] results in distortion D being an exponential function of rate R: D = $Ee^{\alpha R}$, where α is a constant and E is a *function* of the power spectrum density (PSD) of the coefficients. Based on the $i.i.d.$ memoryless source assumption, the classical model is simplified to [6], [12], [14]:

$$D = \varepsilon^2 \sigma_X^2 2^{-2R}, \tag{2}$$

where σ_X^2 denotes signal variance and ε^2 is a source dependent parameter equal to one for uniform distribution, 1.4 for Gaussian distribution, and 1.2 for Laplacian distribution [4]. In reality, few source data are memoryless, and thus some content-dependent heuristic parameters are added in (2) to provide a better modeling of the R-D curve [4], [14].

An alternative approach based on the Laplacian assumption of source data and a Taylor expansion of the classical model (2) is proposed by Chiang *et al.* [1], where rate R is a linear combination of 1/D and 1/D^2:

$$R = aD^{-1} + bD^{-2}, \tag{3}$$

and parameters a, b are obtained from multiple empirical samples of the R-D curve.

Finally, a classical distortion model built for uniform quantizers (UQ) is often used for a variety of sources due to its simplicity [4]:

$$D(\Delta) = \frac{\Delta^2}{\beta}, \tag{4}$$

where β is 12.

To illustrate the accuracy of these models, we plot the R-D curve for frame 0 and frame 252 of Foreman CIF in Fig. 4. From the figures, we observe that a large mismatch exists between these models and the real R-D curve. The mismatch can be explained from two angles. First, all classical models are built on the assumption of a single statistical distribution of the input source data, while the statistical properties of FGS are not accurately modeled by a single distribution function. Second, bitplane coding applied in FGS has specific characteristics that make it different from regular quantizers used in the base layer.

3.2. Bitplane Coding

A video-coding scheme usually has three stages: transform coding, quantization, and then entropy coding [8]. In current image and video coding standards, such as MPEG-2, H.263, and MPEG-4 (base layer), each DCT coefficient is quantized by a different uniform quantizer, and then a run-level coding technique is applied to the quantized DCT coefficients. Since the distortion in run-level coding is controlled by the quantization step size, many studies [4], [5] focus on modeling the distortion function of uniform quantizers.

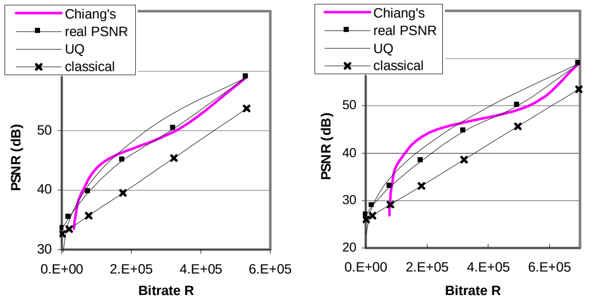

Fig. 4. Classical model (2), Chiang's model (3), real R-D curve, and UQ model (4) for frame 0 of CIF Foreman (left). Same experimental results for frame 252 of CIF Foreman (right).

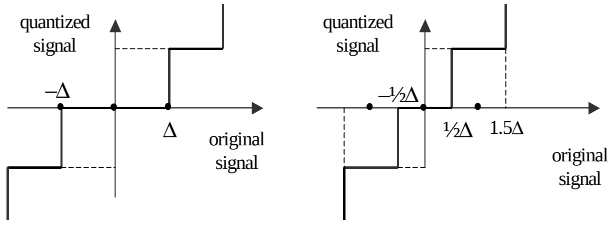

Fig. 5. Bitplane coding (left). Uniform quantizer (right).

FGS uses *bitplane coding*, which considers each input value as a binary number instead of a decimal integer. During FGS streaming, transmission of bitplane i is similar to applying a quantizer with a quantization level Δ equal to $2^{max\_layer-i}$. Although bitplane coding appears to be similar to the uniform quantizer, they are somewhat different. In a uniform quantizer, the reconstruction points are midway between quantization levels [3], while in bitplane coding, the data is reconstructed at exactly[1] the quantization levels themselves as shown in Fig. 5.

3.3. Proposed Model

Let Y be a random variable at the input, distortion D is measured by the Mean Square Error (MSE) that is defined as $E[(Y - \hat{Y})^2]$, where $\hat{Y}$ is a distorted version of Y [2]. For any quantization step Δ, the discrete source MSE function is given by [4]:

$$\begin{aligned} D(\Delta) = &\sum_{k=0}^{N/\Delta} \sum_{n=k\Delta}^{(k+1)\Delta-1} (n - k\Delta)^2 p(n) + \\ &+ \sum_{k=-N/\Delta}^{-1} \sum_{n=k\Delta}^{(k+1)\Delta-1} (n - (k+1)\Delta)^2 p(n), \end{aligned} \tag{5}$$

where N=$2^{max\_layer+1}$.

[1] Note that the MPEG-4 FGS standard allows quarter-point quantizers; however, this option can be turned off and it further does not contribute to the understanding of the rest of the paper. We omit it for clarity.

Since we deal with zero-mean, symmetric exponential data, i.e., $p(n) = ae^{b|n|}$, (5) can be simplified as:

$$\begin{aligned} D(\Delta) &= 2\sum_{k=0}^{N/\Delta}\sum_{i=0}^{\Delta-1}(k\Delta+i-k\Delta)^2 p(k\Delta+i) \\ &= 2a\sum_{k=0}^{N/\Delta} e^{bk\Delta}\sum_{i=0}^{\Delta-1} i^2 e^{bi}. \end{aligned} \qquad (6)$$

Furthermore, since k and i are independent of each other, we compute terms $\sum e^{kb\Delta}$ and $\sum i^2 e^{bi}$ separately. Notice that the first term is a geometric series and can be expanded to:

$$\sum_{k=0}^{N/\Delta} e^{bk\Delta} = \frac{1-e^{b(N+1)\Delta}}{1-e^{b\Delta}} \approx \frac{1}{1-e^{b\Delta}}, \qquad (7)$$

while $\sum i^2 e^{bi}$ can be estimated using integration:

$$\begin{aligned} &\sum_{i=0}^{\Delta-1} i^2 e^{bi} \approx \int_0^{\Delta-1} x^2 e^{bx} dx = \\ &= \tfrac{e^{b(\Delta-1)}}{b}\left[(\Delta-1)^2 - \tfrac{2(\Delta-1)}{b} + \tfrac{2}{b^2}\right] - \tfrac{2}{b^3}. \end{aligned} \qquad (8)$$

Combining (7) and (8), the distortion model $D(\Delta)$ becomes:

$$\begin{aligned} &D(\Delta) = \tfrac{2a}{(1-e^{b\Delta})b}\times \\ &\times\left(e^{b(\Delta-1)}\left[\left(\Delta-1-\tfrac{1}{b}\right)^2+\tfrac{1}{b^2}\right]-\tfrac{2}{b^2}\right). \end{aligned} \qquad (9)$$

For the mixture Laplacian function (1), the final distortion formula is simply a linear combination of two functions in (9). In the first function, $a = p\frac{\lambda_0}{2}$ and $b = -\lambda_0$, while in the second function $a = (1-p)\frac{\lambda_1}{2}$ and $b = -\lambda_1$.

3.4. Experimental Results

To demonstrate the accuracy of classical model (2), UQ model (4), and our model (9), we examine the *average absolute error* (measured in dB and averaged across all bitplanes) of these models in Foreman CIF and Coastguard CIF. The two charts in Fig. 6 show that both the classical and the UQ distortion models are much less accurate in the FGS enhancement layer than the more advanced model examined in this work.

Finally, note that additional experiments using other FGS sequences confirm that our model significantly outperforms the classical and UQ models; however, due to a lack of space, we cannot present these results here.

4. CONCLUSION

This paper posed a question of how well traditional R-D models approximate characteristics of MPEG-4 FGS and possibly other scalable (embedded) coders. We found that much better models can be build if one takes into account the shape of typical PMFs found in real DCT residue. This work proposed a mixture-Laplacian statistical model for DCT residue and derived an accurate closed-form distortion function for such sources. Besides advancing the generic understanding of R-D properties of FGS, this paper also provides a good starting point for further research on FGS streaming. For instance, this simple but efficient distortion model allows servers to implement better congestion control [9] and achieve constant quality in real-time streaming over the Internet.

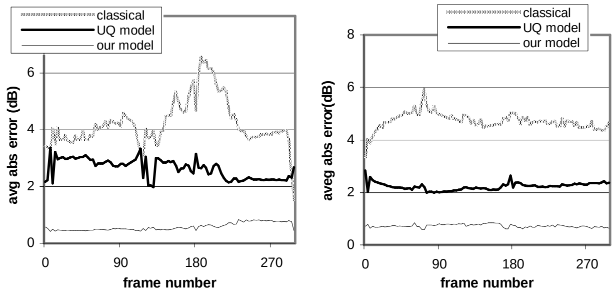

Fig. 6. The average absolute error of classical model (2), UQ model (4), and our model in Foreman CIF (left) and Coastguard CIF (right).

5. REFERENCES

[1] T. Chiang and Y.Q. Zhang, "A new Rate Control Scheme Using Quadratic Distortion Model," *IEEE Trans. CSVT*, vol. 7, Feb. 1997.

[2] T.M. Cover and J.A. Thomas, "Elements of Information Theory," *Wiley, New York, NY*, 1991.

[3] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. on Information Theory*, vol. 44, Oct. 1998.

[4] H.-M. Hang and J.-J. Chen, "Source model for transform video coder and its application. I. Fundamental theory," *IEEE Trans. CSVT*, vol.7, April 1997.

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

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

[7] J. Lin and A. Ortega, "Bit-rate control using piecewise approximation rate-distortion characteristics," *IEEE Trans. CSVT*, vol. 8, Aug. 1998.

[8] F. Ling, W. Li and H. Sun, "Bitplane Coding of DCT Coefficients for Image and Video Compression," *SPIE Conf. on Visual Communications and Image Processing* (VCIP), Jan. 1999.

[9] D. Loguinov and H. Radha, "Open-loop Rate Control for Real-time Video Streaming: Analysis of Binomial Algorithms," *IEEE International Conference on Image Processing (ICIP)*, Sept. 2002.

[10] A. Mallart and F. Falzon, "Analysis of Low Bit Rate Image Transform Coding," *IEEE Trans. Signal Processing*, vol. 46, April 1998.

[11] MPEG, "Information Technology – Coding of Audio Visual Objects – Part 2: Visual AMENDMENT 4: Steaming Video Profile," MPEG 2000/N3518, July 2000.

[12] A.N. Netravali and B.G. Haskell, "Digital Pictures Presentation, Compression, and Standards," *New York, NY: Plenum*, 1988.

[13] H. Radha, M. v. d. Schaar and Y.Chen, "The MPEG-4 fine-grained scalable video coding method for multimedia streaming over IP," *IEEE Trans. Multimedia*, vol. 3, Mar. 2001.

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

[15] L. Zhao, J. Kim, and C.-C. J. Kuo, "MPEG-4 FGS Video Streaming with Constant-Quality Rate Control and Differentiated Forwarding," *VCIP*, 2002.

[16] 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," *IEEE Packet Video*, 2002.