

RATE-DISTORTION EVALUATION FOR TWO-LAYER CODING SYSTEMS

Philippe Hanhart and Touradj Ebrahimi

Multimedia Signal Processing Group, EPFL, Lausanne, Switzerland

ABSTRACT

The Bjøntegaard model is widely used to calculate the compression efficiency between different codecs. However, this model is not sufficient to investigate the impact on quality of the interaction of the base and enhancement layer bit rates when comparing two-layer coding systems. Therefore, in this paper, we propose an extension of the Bjøntegaard model from rate-distortion (R-D) curve fitting to rate-rate-distortion (R^2 -D) surface fitting. The proposed model uses a cubic surface as fitting function and a more complex characterization of the domain formed by the data points to compute a more realistic estimate of compression efficiency. The proposed model can be used to measure the compression efficiency of two-layer coding systems, as well as for other applications, e.g., to optimize the bit rate allocation between texture and depth in 3D video coding. Two examples of assessment of compression efficiency in JPEG XT are presented as illustrations.

Index Terms— Two-layer coding, compression efficiency, coding efficiency, rate-distortion optimization, bit rate allocation, PSNR, Bjøntegaard delta, BD-PSNR

1. INTRODUCTION

In the recent years, layered coding [1] has gained a large popularity in the image and video compression community. Multilayer coding systems partition the information between one base layer and one or more enhancement layers. This approach is typically used for scalable coding, where the enhancement layers can provide spatial, temporal, or quality improvements when compared to the base layer. Additional scalable features, e.g., bit depth, color gamut, or hybrid coding, can also be implemented. Several standards, e.g., JPEG 2000 [2], scalable video coding (SVC) extension of H.264/AVC [3], and scalable H.265/HEVC (SHVC) [4] rely on layered coding to provide scalability. Backward compatibility is another feature that can be implemented using two-layer coding: the base layer is encoded using a legacy encoder for backward compatibility, whereas the enhancement layer is encoded using a different and optimized coding

scheme. JPEG XT [5] and multi-resolution frame-compatible (MFC) stereo coding [6] are examples of backward compatible standards using two-layer coding.

Quality assessment of images and video sequences is generally performed through objective metrics or subjective tests. Whereas subjective tests are undeniably the most accurate means to evaluate quality, as measurements are performed by human observers, they are time consuming, expensive, and not always feasible. In practice, objective metrics are often a more preferable and more efficient alternative, even if they do not accurately reflect human perception of visual quality [7]. For codec optimization, where several parameters can be tuned to improve quality, subjective evaluations are impractical. However, previous studies [8, 9] have shown that the peak signal-to-noise ratio (PSNR) metric, which is commonly used by coding experts, is a reliable metric as long as the content is not changed.

To calculate the compression efficiency between different codecs based on PSNR measurements, the Bjøntegaard model [10] is typically used. This model computes the average PSNR and bit rate differences between two rate-distortion (R-D) curves obtained from the PSNR measurements when encoding a content at different bit rates. Xiang *et al.* [11] have proposed an extension of the Bjøntegaard model, referred to as generalized BD-PSNR, to take coding complexity into account. In a previous work [12], we have also proposed a model to calculate the average coding efficiency based on subjective quality scores. However, none of these models can be used to investigate the impact on quality of the interaction of the base and enhancement layers bit rates when comparing two-layer coding systems. Therefore, in this paper, we propose an extension of the Bjøntegaard model from R-D curve fitting to rate-rate-distortion (R^2 -D) surface fitting. The proposed model uses a cubic surface as fitting function. While the generalized BD-PSNR model [11] only considers a rectangular domain in the RC -plane to evaluate the delta PSNR, the proposed model uses a more complex characterization of the domain formed by the data points to compute a more realistic estimate of the compression efficiency.

The remainder of this paper is organized as follows. A brief overview of the Bjøntegaard model is given in Sec. 2. The proposed model is described in details in Sec. 3. Examples of application and discussions are presented in Sec. 4. Finally, concluding remarks are given in Sec. 5.

This work has been conducted in the framework of the Swiss National Foundation for Scientific Research (FN 200021-143696-1), the Swiss SERI (C12.0081), and COST IC1005 The digital capture, storage, transmission and display of real-world lighting HDRi.

2. BJØNTEGAARD MODEL

The Bjøntegaard model [10] measures the compression efficiency between two different coding algorithms. To approximate a rate-distortion (R-D) curve, a third order logarithmic polynomial fitting has been proposed in the Bjøntegaard model, based on experimental observations

$$\hat{D}(R) = a \log^3 R + b \log^2 R + c \log R + d \quad (1)$$

where \hat{D} is the fitted distortion in PSNR, R is the bit rate, and a , b , c , and d are the parameters of the fitting function.

To simplify notation, in the rest of the paper, we use lower case r when referring to the logarithm of the bit rate, i.e., $r = \log R$. Therefore, Eq. (1) is rewritten as

$$\hat{D}(r) = ar^3 + br^2 + cr + d \quad (2)$$

At least four (R, D) pairs are required to determine the fitting parameters of Eq. (2). If more than four pairs are used, then the R-D values are fitted in a least square sense.

The average PSNR difference between two R-D curves is approximated by the difference between the integrals of the fitted R-D curves divided by the integration interval [10]

$$\Delta D = E[D_2 - D_1] \approx \frac{1}{r_H - r_L} \int_{r_L}^{r_H} [\hat{D}_2(r) - \hat{D}_1(r)] dr \quad (3)$$

where ΔD is the so-called Bjøntegaard delta PSNR (BD-PSNR) computed between the two fitted R-D curves $\hat{D}_1(r)$ and $\hat{D}_2(r)$, respectively, and the integration bounds, r_L and r_H , are determined by the range of bit rate values common to both fitted R-D curves to avoid extrapolation.

To express the (logarithm of the) rate as a function of the distortion, a third order polynomial fitting has been proposed in the Bjøntegaard model to fit the R-D values

$$\hat{r}(D) = aD^3 + bD^2 + cD + d \quad (4)$$

Note that a second fitting process is required to fit the bit rate values and that Eq. (4) is not the inverse function of Eq. (2).

The average bit rate difference between two R-D curves is approximated as [10]

$$\begin{aligned} \Delta R &= E \left[\frac{R_2 - R_1}{R_1} \right] = E \left[\frac{R_2}{R_1} \right] - 1 = E [10^{r_2 - r_1}] - 1 \\ &\approx 10^{E[r_2 - r_1]} - 1 \approx 10^{\frac{1}{D_H - D_L} \int_{D_L}^{D_H} [\hat{r}_2(D) - \hat{r}_1(D)] dD} - 1 \end{aligned} \quad (5)$$

where ΔR is the so-called Bjøntegaard delta rate (BD-Rate) computed between the two fitted R-D curves $\hat{r}_1(D)$ and $\hat{r}_2(D)$, respectively, and the integration bounds, D_L and D_H , are determined by the range of PSNR values common to both fitted R-D curves to avoid extrapolation.

Thanks to the logarithmic bit rate scale, the estimation of the average bit rate reduction is also simplified.

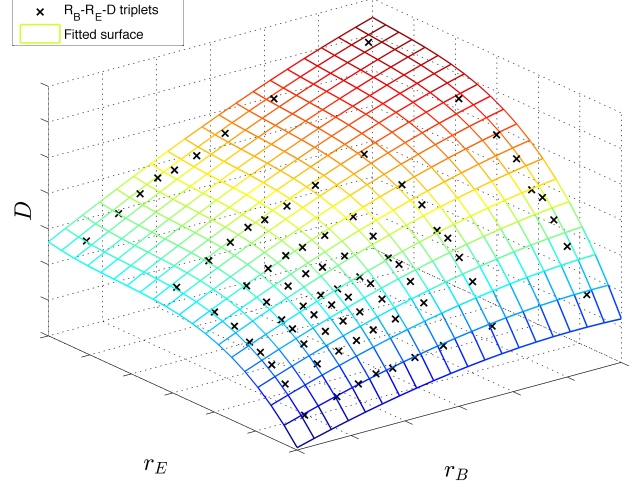


Fig. 1: R_B - R_E - D surface fitting.

3. PROPOSED MODEL

In this section, we propose an extension of the Bjøntegaard model for measuring the compression efficiency between two rate-rate-distortion (R^2 - D) surfaces. First, the function used to fit the R^2 - D surfaces is described. Then, the calculation of average PSNR and bit rate differences between two fitted R^2 - D surfaces is presented. A MATLAB implementation of the proposed model can be downloaded from: <http://mmspg.epfl.ch/2dbd>.

3.1. Fitting Function

The Bjøntegaard model uses a cubic function to fit the R-D curve, based on the observation that R-D values expressed in $(\log(\text{bit rate}), \text{PSNR})$ do not deviate much from straight lines [13]. Following the same principle, we propose to use a cubic surface to fit the R^2 - D surface. The cubic surface is given by

$$z(x, y) = \sum_{(i,j) \in S} p_{ij} x^i y^j \quad S = \{(i, j) \in \mathbb{N}^2 | i+j \leq 3\} \quad (6)$$

where p_{ij} are the parameters of the fitting function.

The cross-terms, i.e., $p_{11}xy$, $p_{21}x^2y$, and $p_{12}xy^2$, allow more flexibility for the fitting of the R^2 - D surface, which improves the goodness of the fit, but increases the number of required data points. At least ten (x, y, z) triplets are required to determine the fitting parameters of Eq. (6). If more than ten triplets are used, then the data points are fitted in a least square sense. However, in practice, to obtain a more realistic estimate of the performance evaluation, 16 or more triplets should be used. Fig. 1 shows the fitting result for one HDR image encoded with JPEG XT. As it can be observed, the fitting accuracy is quite good.

3.2. Average PSNR Difference

The R²-D surface is obtained by varying one parameter of the base and enhancement layers in coding schemes while measuring the PSNR of the reconstructed image or video sequence. Considering M base layer parameter values $(P_{B,1}, \dots, P_{B,M})$ and N enhancement layer parameter values $(P_{E,1}, \dots, P_{E,N})$, this yields to a set of $M \times N$ base layer bit rate values $(R_{B,11}, \dots, R_{B,MN})$ and enhancement layer bit rate values $(R_{E,11}, \dots, R_{E,MN})$ with corresponding PSNR values (D_{11}, \dots, D_{MN}) . The corresponding R²-D surface is fitted in a least square sense using a cubic surface

$$\hat{D}(r_B, r_E) = \sum_{(i,j) \in S} p_{ij} r_B^i r_E^j \quad S = \{(i,j) \in \mathbb{N}^2 | i+j \leq 3\} \quad (7)$$

where \hat{D} is the fitted distortion in PSNR, r_B and r_E are the logarithms of the base and enhancement layers bit rates, respectively, and p_{ij} are the parameters of the fitting function.

Similarly to the Bjøntegaard model [10], the average PSNR difference between two R²-D surfaces is approximated by the difference between the integrals of the fitted R²-D surfaces divided by the area of the integration domain

$$\begin{aligned} \Delta D &= E[D_2 - D_1] \\ &\approx \frac{1}{|A|} \iint_A [\hat{D}_2(r_B, r_E) - \hat{D}_1(r_B, r_E)] dr_B dr_E \quad (8) \end{aligned}$$

where ΔD is the delta PSNR computed between the two fitted R²-D surfaces $\hat{D}_1(r_B, r_E)$ and $\hat{D}_2(r_B, r_E)$, respectively, and the integration domain A is given by the intersection of the domains on which the two R²-D surfaces are fitted

$$A = A_1 \cap A_2 \quad (9)$$

Fig. 2 illustrates the (r_B, r_E, D) triplets projected on the $r_B r_E$ -plane. The data points form a domain defined by four corners corresponding to the extrema of P_B and P_E (see Fig. 2). The domain is delimited by the four contours connecting the four corners. The contour which starts at I and ends at J is defined by the pairs $((r_{B,11}, r_{E,11}), \dots, (r_{B,M1}, r_{E,M1}))$. We propose to fit these pairs with a cubic curve to estimate the contour

$$\hat{r}_E(r_B) = ar_B^3 + br_B^2 + cr_B + d \quad (10)$$

The same principle is applied to estimate the three remaining contours, with the exception that the contours between I and K and between J and L are expressed as a function of r_E .

The domain delimited by the four contours (represented in gray in Fig. 2) is thus defined as

$$A = \{(r_B, r_E) \in \mathbb{R}^2 | \beta_{\min}(r_E) \leq r_B \leq \beta_{\max}(r_E), \epsilon_{\min}(r_B) \leq r_E \leq \epsilon_{\max}(r_B)\} \quad (11)$$

where the functions β and ϵ are extensions of the contour fitting functions that simply perform repetition for points that lie

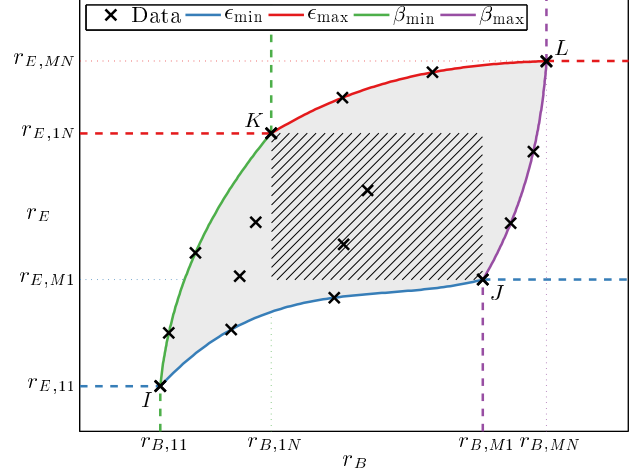


Fig. 2: Domain on which the R²-D surface is fitted. The hatched area represents a simple integration domain based on min and max values, as used in [11], while the proposed model integrates over the whole area (represented in gray).

outside the range of fitted values (as illustrated by the dashed lines in Fig. 2). For example, ϵ_{\min} is the extension of the contour which starts at I and ends at J

$$\epsilon_{\min}(r_B) = \begin{cases} \hat{r}_E(r_{B,11}) & \text{if } r_B < r_{B,11} \\ \hat{r}_E(r_{B,M1}) & \text{if } r_B > r_{B,M1} \\ \hat{r}_E(r_B) & \text{otherwise} \end{cases} \quad (12)$$

The same principle is used for the other extensions.

The domain on which the R²-D surface is fitted is determined independently for both surfaces following the procedure described here above. Then, the integral is evaluated on the intersection of the two domains. Even though the analytical form of the integral can be easily determined, its evaluation would require a complex parametrization of the integration bounds. Therefore, the integral is approximated using a finite sum. Note that in the generalized BD-PSNR [11] model, the integration domain corresponds to a rectangular domain defined by the extreme values (as represented by the hatched area in Fig. 2). However, this simple integration domain might not be representative of the full domain.

3.3. Average Bit Rate Difference

The computation of the average base and enhancement layers bit rate differences between two R²-D surfaces is performed following a similar procedure as for the average PSNR difference. First, a new fitting process is performed using the cubic surface. Note that the inverse function of a cubic function can be determined using Cardano's formula, but fitting a new surface yields better accuracy. Then, the delta base layer rate, ΔR_B , and delta enhancement layer rate, ΔR_E , are computed on the integration domain, which is determined following a similar procedure as in Sec. 3.2.

Table 1: Average coding efficiency of the JPEG XT HDR image compression standard.

(a) Average coding efficiency of the three main profiles, computed over 17 HDR images and 5 TMOs.

Profile	ΔD relative to (dB)			ΔR_B relative to (%)			ΔR_E relative to (%)		
	A	B	C	A	B	C	A	B	C
A	-	-0.84	-3.86	-	-7.46	+52.40	-	+3.23	+66.15
B	+0.84	-	-3.24	+14.51	-	+84.25	+3.27	-	+30.94
C	+3.86	+3.24	-	-29.42	-32.88	-	-30.78	-18.34	-

(b) Influence of the TMO on the coding efficiency of profile B.

TMO	ΔD relative to (dB)					ΔR_B relative to (%)					ΔR_E relative to (%)				
	d03	γ	m11	m06	r02	d03	γ	m11	m06	r02	d03	γ	m11	m06	r02
<i>drago03 (d03)</i>	-	+2.8	-2.3	-3.4	+1.1	-	-24.4	+19.3	+34.9	-16.7	-	-32.3	+3.5	+66.4	-13.2
<i>gamma (γ)</i>	-2.8	-	-4.7	-6.2	-1.7	+35.0	-	+73.8	+80.1	+16.8	+58.6	-	+30.8	+101.4	+44.6
<i>mai11 (m11)</i>	+2.3	+4.7	-	-1.2	+3.2	-13.7	-37.8	-	+23.7	-30.2	+0.8	-22.0	-	+44.4	-3.2
<i>mantiuk06 (m06)</i>	+3.4	+6.2	+1.2	-	+4.6	-19.4	-36.8	-11.6	-	-27.7	-31.9	-46.1	-28.6	-	-34.2
<i>reinhard02 (r02)</i>	-1.1	+1.7	-3.2	-4.6	-	+21.5	-13.0	+55.2	+47.1	-	+16.4	-25.5	+6.7	+59.6	-

A negative (positive) delta PSNR ΔD indicates a decrease (increase) of PSNR for the same base and enhancement layer bit rates. A negative (positive) delta base layer rate ΔR_B indicates a decrease (increase) of the base layer bit rate for the same PSNR and enhancement layer bit rate. The same principle applies to the delta enhancement layer rate ΔR_E .

4. APPLICATIONS AND DISCUSSIONS

In this section, we show two examples of application of the proposed model. For this purpose, we used a dataset of 17 high dynamic range (HDR) image, which were encoded with JPEG XT [5]. JPEG XT is based on a two-layer design and is backward compatible with the popular JPEG compression standard. The base layer contains a low dynamic range (LDR) image, which is a tone-mapped version of the HDR image, accessible to legacy implementations, while the enhancement layer allows recovering the full dynamic range. The three main profiles of JPEG XT were used. For each profile, the quality factor of the base and enhancement layers was varied in the range [20, 98] with a step of 2. Five different tone-mapping operators (TMOs) were considered to create the base layer LDR image. The dataset consisted of 17 HDR images \times 3 profiles \times 40 base layer quality values \times 40 enhancement layer quality values \times 5 TMOs = 408,000 compressed images with corresponding PSNR values.

Table 1 (a) reports the average coding efficiency of the three main profiles of JPEG XT. The values were averaged over the 17 HDR images and 5 TMOs. Results show that profile C provides a gain of over 3.2dB in PSNR for the same bit rate when compared to profiles A and B, whereas profile B provides a gain of about 0.84dB over profile A. On the other hand, for the same PSNR and enhancement layer bit rate, the bit rate of the base layer can be reduced by about 7.46% for profile A when compared to profile B. For the same base layer bit rate, i.e., same quality of the LDR image, the enhancement layer bit rate for profile C can be reduced by about 30% and 20% when compared to profiles A and B, respectively.

Table 1 (b) reports the influence of the TMO on the coding efficiency of profile B. The values were averaged over the 17 HDR images. Surprisingly, the simple *gamma* TMO, which is very easy to inverse to predict the HDR image from the LDR image, reduces the PSNR of the reconstructed HDR image by 1.7dB to 6.2dB for the same base layer and en-

hancement layer rates. On the other hand, the *mantiuk06* TMO, which usually produces pleasant LDR images, allows reducing the bit rate of the enhancement layer by 28% to 46% when compared to other TMOs.

As it can be observed, the proposed model allows a more complete and detailed quantitative analysis when compared to the analysis reported by Pinheiro *et al.* [14]. The results reported in [14] are more qualitative and their analysis was mostly performed on two dimensions only (distortion and enhancement layer bit rate), by fixing the quality parameter of the base layer. Additionally, the proposed model can be used for other applications than two-layer coding. For example, this model can be used for video plus depth or multiview video plus depth coding, to find the optimal bit rate allocation between the texture and depth data, to maximize the quality of a synthesized viewpoint.

Note that the proposed model only considers one distortion, e.g., the distortion of the base or residual layer, or the distortion of a derived image/video sequence (see example above). To consider two different distortions, e.g., the base and enhancement layer distortions, a 4D model must be used. In this case, 20 or more quads are required, while most performance analysis are conducted with only 4×4 combinations of base and enhancement layer parameter values.

5. CONCLUSION

In this paper, we proposed an extension of the Bjøntegaard model from rate-distortion curve fitting to rate-rate-distortion surface fitting. The proposed model uses a cubic surface as fitting function and a more complex characterization of the domain formed by the data points to compute a more realistic estimate of the compression efficiency. We presented two applications of the proposed model to measure the compression efficiency of JPEG XT. The proposed model can also be used for other applications, e.g., to optimize the bit rate allocation between texture and depth in 3D video coding.

6. REFERENCES

- [1] M. Ghanbari, "Two-layer coding of video signals for VBR networks," *IEEE Journal on Selected Areas in Communications*, vol. 7, no. 5, pp. 771–781, June 1989.
- [2] A. Skodras, C. Christopoulos, and T. Ebrahimi, "The JPEG 2000 still image compression standard," *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp. 36–58, Sept. 2001.
- [3] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1103–1120, Sept. 2007.
- [4] Y. Ye and P. Andrivon, "The Scalable Extensions of HEVC for Ultra-High-Definition Video Delivery," *IEEE MultiMedia*, vol. 21, no. 3, pp. 58–64, July 2014.
- [5] T. Richter, "On the standardization of the JPEG XT image compression," in *Picture Coding Symposium (PCS)*, Dec. 2013.
- [6] T. Lu, H. Ganapathy, G. Lakshminarayanan, T. Chen, W. Husak, and P. Yin, "Orthogonal Muxing Frame Compatible Full Resolution technology for multi-resolution frame-compatible stereo coding," in *IEEE International Conference on Multimedia and Expo (ICME)*, July 2013.
- [7] H. Sheikh, M. Sabir, and A. Bovik, "A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.
- [8] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," *Electronics Letters*, vol. 44, no. 13, pp. 800–801, June 2008.
- [9] J. Korhonen and J. You, "Peak signal-to-noise ratio revisited: Is simple beautiful?" in *International Workshop on Quality of Multimedia Experience (QoMEX)*, July 2012.
- [10] G. Bjøntegaard, "Calculation of average PSNR differences between RD-curves," ITU-T SG16/Q6, Austin, Texas, USA, Tech. Rep. VCEG-M33, Apr. 2001.
- [11] X. Li, M. Wien, and J.-R. Ohm, "Rate-complexity-distortion evaluation for hybrid video coding," in *IEEE International Conference on Multimedia and Expo (ICME)*, July 2010.
- [12] P. Hanhart and T. Ebrahimi, "Calculation of average coding efficiency based on subjective quality scores," *Journal of Visual Communication and Image Representation*, vol. 25, no. 3, pp. 555–564, 2014.
- [13] G. Bjøntegaard, "Improvements of the BD-PSNR model," ITU-T SG16/Q6, Berlin, Germany, Tech. Rep. VCEG-AI11, July 2008.
- [14] A. Pinheiro, K. Fliegel, P. Korshunov, L. Krasula, M. Bernardo, M. Pereira, and T. Ebrahimi, "Performance evaluation of the emerging JPEG XT image compression standard," in *International Workshop on Multimedia Signal Processing (MMSP)*, Sept. 2014.