SEQUENCE-LEVEL MODELS FOR DISTORTION-RATE BEHAVIOUR OF COMPRESSED VIDEO

Lai U Choi[1], *Michel T. Ivrlač*[2], *Eckehard Steinbach*[1], *and Josef A. Nossek*[2]

[1]Institute of Communication Networks
Media Technology Group
Munich University of Technology
email: {laiuchoi, eckehard.steinbach}@tum.de

[2]Institute for Circuit Theory and Signal Processing
Wireless Communications
Munich University of Technology
email: {ivrlac, josef.a.nossek}@tum.de

ABSTRACT

In this paper[1], two empirical models for the sequence-level distortion-rate performance of predictive video source encoding are proposed. They require very limited amount of empirical data, namely three pairs of rate and distortion, in order to set up the model parameters. The advantages of these proposed models are the robustness towards measurement noise, low number of model parameters with a closed form setup solution, and high accuracy. Experimental validation using H.264/AVC encoded video test sequences reports high accuracy of these two proposed models.

1. INTRODUCTION

Optimization of so-called end-to-end quality of the transmission of compressed video has recently gained attention in the research community, both for the wireline Internet and for its wireless extensions [1, 2, 3, 4]. Due to its conceptual simplicity and mathematical convenience, it is common to associate video quality with the average pixel-by-pixel distortion of the video frames. This distortion results from both the effects of lossy compression introduced by source encoding, and the lossy transmission channel. Accurate modeling of the different types of distortion is the foundation to successful end-to-end quality optimization.

In this paper, two models are developed for the sequence-level distortion-rate (D-R) performance of predictive video source encoding. Both models are of empirical nature, treating the source encoder as a black box, focusing on its input-output behaviour. Both of them require very limited amount of empirical data, namely three pairs of rate and distortion, in order to set up the model parameters. The parameter setup procedure is derived in closed form without need of non-linear optimization. Experimental validation using H.264/AVC encoded video test sequences reports high accuracy of these two proposed models.

Most of the D-R models proposed in literature (e.g. [5, 6, 7]) are designed for a single frame or macroblock and are usually used for rate-control. Other approaches using rate-distortion and quantization theory (e.g. [8, 9]), usually fail to describe the behaviour of a real video encoder with enough accuracy. Although some sequence-level D-R models have been proposed [1, 2], there is still room to improve the accuracy and the practical applicability of the models. For instance, the D-R model from [1] is sensitive to measurement noise and may lead to rather surprising results, like negative values of predicted mean square error distortion. The advantages of our proposed D-R models are the robustness towards measurement noise, low number of model parameters with a closed form setup solution, and high accuracy. The robustness property is due to the structure of the models which guarantees convexity and monotonicity of the D-R function, and positive values of mean square error distortion.

In the following, we will develop the two proposed models in detail. The first model (see Section 3) is based on the MSE, while the second (see Section 4) relies on the PSNR as the measure of distortion. Experimental validation for both models is provided in Section 5.

2. DISTORTION MEASURE

It is common to base the measure of distortion of a video sequence on the mean square error (MSE) between the sequence of original and reconstructed frames:

$$\mathrm{MSE} = \frac{1}{NMF} \sum_{f=1}^{F} \sum_{n=1}^{N} \sum_{m=1}^{M} \left(Y_f(n,m) - Y'_f(n,m) \right)^2 , \tag{1}$$

where $Y_f(n,m)$ and $Y'_f(n,m)$ is the luminance of the pixels at the position (n,m) of the f-th original and reconstructed frame, respectively. The averaging of the squared error takes place over F consecutive frames of resolution of $N \times M$ pixels. It is also common to specify the distortion in terms of the *peak signal to noise ratio* (PSNR), which is based on the MSE:

$$\mathrm{PSNR} \;=\; 10 \log_{10} \frac{255^2}{\mathrm{MSE}}. \tag{2}$$

The source distortion depends on the data rate R that is generated by the source encoder. The relationship between the

[1]This work is supported partially by an Alexander von Humboldt (AvH) Research Fellowship.

0-7803-9134-9/05/$20.00 ©2005 IEEE

rate and the distortion is defined by the so-called *distortion-rate* (D-R) function.

3. MSE-BASED D-R MODEL

Let us have a look at the first proposed model. We propose to model the relationship between source rate R and the distortion MSE by the following simple equation:

$$\mathrm{MSE}(R) = \frac{a}{\exp\left(R/b\right) - 1}, \tag{3}$$

where a and b are two model coefficients which have to be fitted to measured distortion data. As will be demonstrated, (3) constitutes a highly accurate model for the MSE rate distortion of a H.264/AVC compressed video sequence. Based on only *three* measured pairs of rate and distortion, a highly accurate prediction of distortion at other rates can be obtained. Let us now determine the coefficients a and b. With the definitions

$$x = \exp(R/b), \quad \text{and} \quad y = 1/\mathrm{MSE}, \tag{4}$$

it follows from (3) that the proposed D-R function requests a *linear* relationship between x and y:

$$y = \left(x - 1\right)/a. \tag{5}$$

There is experimental evidence that by proper choice of the coefficient b this linear relationship is indeed fulfilled fairly well by video sequences compressed by the H.264/AVC standard. The coefficient b can be determined analytically from only *three* pairs (R_i, MSE_i) of source rate and distortion, where $i \in \{1, 2, 3\}$. Calling (x_i, y_i) the corresponding values obtained from (4), we need to have

$$\frac{x_3 - x_2}{x_2 - x_1} = \frac{y_3 - y_2}{y_2 - y_1}, \tag{6}$$

in order to have all three pairs (x_i, y_i) aligned on a straight line. With (6) and (4) then follows

$$\gamma \cdot \exp\left(\frac{R_1 - R_2}{b}\right) + \exp\left(\frac{R_3 - R_2}{b}\right) = 1 + \gamma, \tag{7}$$

where γ is defined as

$$\gamma = \frac{\mathrm{MSE}_2/\mathrm{MSE}_3 - 1}{1 - \mathrm{MSE}_2/\mathrm{MSE}_1}. \tag{8}$$

In order to solve (7) analytically for b, we request the following relationship between the rates:

$$R_2 = \frac{R_1 + R_3}{2}, \tag{9}$$

which leads to

$$\Delta R = (R_3 - R_2) = -(R_1 - R_2). \tag{10}$$

Without loss of generality, we can assume $R_3 > R_1$ which leads to $\Delta R > 0$. With the symbol $w = \exp(\Delta R/b)$, we obtain from (7) the quadratic equation:

$$w^2 - (1 + \gamma)w + \gamma = 0. \tag{11}$$

Its two solutions are given by $w \in \{1, \gamma\}$. Since $w = 1$ implies $\Delta R = 0$ for finite b, which is not valid, the only relevant solution is $w = \gamma$. With (8), we therefore obtain for the coefficient b the expression:

$$b = \frac{\Delta R}{\log_{\mathrm{e}} \dfrac{\mathrm{MSE}_2/\mathrm{MSE}_3 - 1}{1 - \mathrm{MSE}_2/\mathrm{MSE}_1}}. \tag{12}$$

After knowing b, the coefficient a can be obtained easily from (3) as:

$$a = \left(\exp(R_1/b) - 1\right)\mathrm{MSE}_1. \tag{13}$$

In case that the requirement (9) is only fulfilled approximately, one can use a modified definition of ΔR. We propose to use the logarithmic mean:

$$\Delta R = \left(\frac{1}{2}(R_2 - R_1)(R_3 - R_2)(R_3 - R_1)\right)^{1/3}, \tag{14}$$

where $(.)^{1/3}$ takes the real third root of its argument. Similarly, we propose to use a logarithmic mean to obtain a as:

$$a = \left(\prod_{i=1}^{3} \left(\exp\left(R_i/b\right) - 1\right)\mathrm{MSE}_i\right)^{1/3}. \tag{15}$$

Note that if (9) is fulfilled exactly, the expressions in (14) and (15) equal the expressions in (10) and (13), respectively. In this way, we have obtained analytical expressions for the model coefficients a and b which assume three pairs of rate and distortion obtained from measurement. To this end, it is best to choose R_1 to be the smallest rate and R_3 to be the largest rate of interest. We will later see that this ensures an accurate prediction of the distortion at any rate in between. It is interesting to note that the D-R function from (3) has the inverse

$$R = b \cdot \log_{\mathrm{e}}\left(1 + \frac{a}{\mathrm{MSE}}\right), \tag{16}$$

which has the same form as the channel capacity of a AWGN channel. This may perhaps lead to interpretation of a and b as sort of power and bandwidth equivalents. Finally, note that both a and b are always positive and the D-R function from (3) is strictly monotonic and convex for positive R. Moreover, the MSE value always remains positive. These properties result in robustness of the proposed D-R model towards mesurement noise and decent behaviour even when used for extrapolation (outside the range of measured source rates).

4. PSNR-BASED D-R MODEL

In the following, we will develop the second proposed D-R function which is based on the PSNR, instead of the MSE, as the measure of distortion. We propose the generic form:

$$\mathrm{PSNR}(R_1) = \mathrm{PSNR}(R_2) \; + \; \mathcal{F}\left(R_1, R_2\right). \qquad (17)$$

In this way, the source distortion at a rate R_1 is based on the source distortion at another rate R_2. An additive correction term is applied in the form of the function $\mathcal{F}\left(R_1, R_2\right)$, of two arguments which returns a real value. Let us now specify the function $\mathcal{F}(R_1, R_2)$. To this end, note that

$$\begin{aligned} \mathrm{PSNR}(R_2) \; &= \; \mathrm{PSNR}(R_1) + \mathcal{F}\left(R_2, R_1\right) \\ &= \; \mathrm{PSNR}(R_2) + \mathcal{F}\left(R_1, R_2\right) + \mathcal{F}\left(R_2, R_1\right). \end{aligned}$$

It therefore follows that $\mathcal{F}\left(R_1, R_2\right) = -\mathcal{F}\left(R_2, R_1\right)$ must hold. As higher rate means higher PSNR, we also need to have $F\left(R_1, R_2\right) > 0$ iff $R_1 > R_2$. Finally, $\mathcal{F}\left(R_1, R_2\right) = 0$ must hold iff $R_1 = R_2$. In summary, the following constraints have to be fulfilled:

$$\mathcal{F}\left(R_1, R_2\right) \begin{cases} = 0 & \text{for } \; R_1 = R_2 \\ > 0 & \text{for } \; R_1 > R_2 \\ = -F\left(R_2, R_1\right) & \end{cases} . \qquad (18)$$

Using the construction

$$\mathcal{F}\left(R_1, R_2\right) = f\left(R_1, R_2\right) - f\left(R_2, R_1\right), \qquad (19)$$

the constraints (18) are met if

$$\left(R_1 > R_2\right) \; \rightarrow \; \left(f(R_1, R_2) > f(R_2, R_1)\right). \qquad (20)$$

There are many different functions f which fullfil (20). However, it turns out that an excellent match to the input-output behaviour of an H.264/AVC compliant or similar source encoder can be achieved using the rather simple function:

$$f(R_1, R_2) \; = \; b \cdot \sqrt{\frac{R_1}{R_2}}, \quad \text{with} \quad b > 0. \qquad (21)$$

Substituting (21) into (19) and the latter into (17) we obtain:

$$\mathrm{PSNR}(R_1) \; = \; \mathrm{PSNR}(R_2) + b \cdot \sqrt{\frac{R_1}{R_2}} \left(1 - \frac{R_2}{R_1}\right). \qquad (22)$$

When (22) is fitted to measured data, it is possible to optimize both the parameter b and the value for the reference rate R_2. When this is done, the PSNR at the reference rate should be part of the optimization, too. In this way, the proposed distortion-rate function is given by:

$$\mathrm{PSNR}(R) \; = \; a \; + \; b \; \cdot \sqrt{\frac{R}{c}} \left(1 - \frac{c}{R}\right), \qquad (23)$$

where the tuple $(a, b, c) \in \Re \times \Re_+^2$ constitutes three coefficients which can be fitted to measured distortion. Herein the symbols $\Re$ and $\Re_+$ denote the sets of real numbers, and positive real numbers, respectively. Note that there is a unique inverse of the D-R function from (23), which is given by

$$R \;\; = \;\; c \cdot \left(z + \sqrt{1 + z^2}\right)^2, \quad \text{where} \qquad (24)$$

$$z \;\; = \;\; \frac{\mathrm{PSNR} - a}{2b}. \qquad (25)$$

By measuring the source distortion PSNR_i at *three* pairwise different source rates R_i, where $i \in \{1, 2, 3\}$, we can determine the coefficients a, b and c in closed form as:

$$c \;\; = \;\; R_1 \cdot \frac{\left(1 - \nu_{2,1}\right)\mu + \nu_{3,1} - 1}{\left(1 - \nu_{1,2}\right)\mu + \nu_{1,3} - 1} \qquad (26)$$

$$b \;\; = \;\; \frac{\mathrm{PSNR}_2 - \mathrm{PSNR}_1}{\xi_2 - \xi_1} \qquad (27)$$

$$a \;\; = \;\; \mathrm{PSNR}_1 - b \cdot \xi_1, \qquad (28)$$

where

$$\nu_{i,j} \;\; = \;\; \sqrt{R_i / R_j}, \qquad (29)$$

$$\xi_i \;\; = \;\; \sqrt{\frac{R_i}{c}} \left(1 - \frac{c}{R_i}\right), \quad \text{and} \qquad (30)$$

$$\mu \;\; = \;\; \frac{\mathrm{PSNR}_1 - \mathrm{PSNR}_3}{\mathrm{PSNR}_1 - \mathrm{PSNR}_2}. \qquad (31)$$

In this way, we have obtained analytical expressions for the model coefficients a, b and c, which assume three pairs of rate and distortion obtained from measurement. Note that the D-R function from (23) is strictly monotonic and convex in R and the corresponding MSE value is always positive. Similarly to the MSE-based D-R model from Section 3, these properties lead to robustness against measurement noise and decent behaviour of the model.

5. EMPIRICAL EVALUATION

The proposed MSE- and PSNR-based D-R models from (3) and (23), respectively, are validated by comparison to measured data generated from H.264/AVC encoded video test sequences ("Foreman" and "Mother & Daughter"). Figure 1 shows the resulting D-R functions (solid curves) together with the measured data (circles). The crosses indicate the three pairs of rate and distortion which are used to set up the model parameters. The distortion is displayed in terms of PSNR for both models. Close observation of Figure 1 reveals that both models deliver a highly accurate prediction of the distortion for a huge range of source rates, starting from 40 kbps and going up to about 2 Mbps (1:50). The root mean square (RMS) and maximum absolute errors in

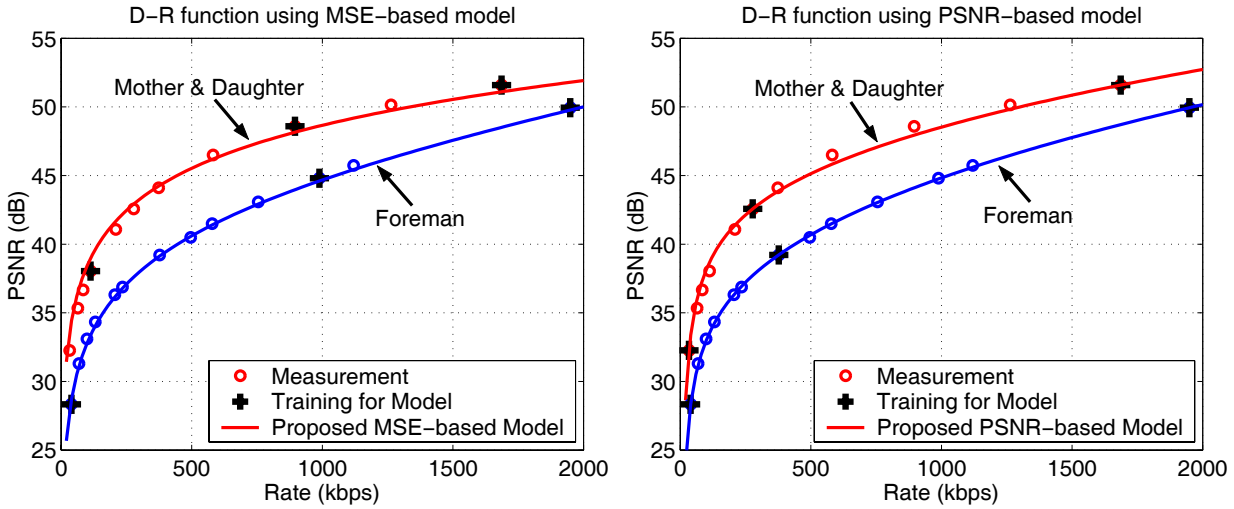

Fig. 1. D-R function of MSE-based model (left) and PSNR-based model (right).

PSNR prediction are shown in Table 1, where the maximum absolute error is displayed in paranthesis. Both models perform excellent for the "Foreman" sequence. The PSNR-based model achieves a maximum absolute error of only 0.145 dB (3.4% maximum error in terms of MSE), while the MSE-based model leads to a maximum error of 8.9% in terms of MSE (0.37 dB). Both models perform worse for the sequence "Mother & Daughter", yet they appear to be still fairly accurate, keeping the maximum absolute error below 1 dB. When the range of supported rates is reduced to $(40 - 400)$ kbps, the accuracy improves to 0.12 dB for the PSNR-based model, as is detailed in Table 2. Note that in our experiment, for the MSE-based model the rate requirement (9) is only fulfilled approximately due to limited experimental data obtained in the test. For instance, in the "Foreman" sequence (full rate range) we use $R_3 - R_2 = 960$ kbps while $R_2 - R_1 = 949$ kbps. We therefore have to resort to (14) and (15) to obtain the model coefficients a and b. It can be expected that that MSE-based model will perform better when the rate requirement (9) is exactly fulfilled.

Table 1. RMS and maximum absolute error in dB for full rate range $(40 - 2000)$ kbps.

Sequence	MSE-based	PSNR-based
Foreman	0.21 (0.37)	0.074 (0.145)
Mother & Daugther	0.51 (0.93)	0.48 (0.70)

Table 2. RMS and maximum absolute error in dB for reduced rate range $(40 - 400)$ kbps.

Sequence	MSE-based	PSNR-based
Foreman	0.20 (0.35)	0.05 (0.11)
Mother & Daugther	0.23 (0.39)	0.069 (0.12)

6. CONCLUSION

Two empirical models for the sequence-level D-R performance of predictive video source encoding have been proposed. The first model is based on the MSE while the second uses PSNR as the measure of distortion. Closed form setup solutions have been derived for both models. The special structure and the low number of model parameters lead to robustness against measurement noise and high model accuracy, which is confirmed by experimental validation using H.264/AVC encoded video test sequences.

7. ACKNOWLEDGEMENT

The authors would like to express their sincere thanks to the Alexander von Humboldt (AvH) foundations for kindly supporting this research.

8. REFERENCES

[1] K. Stuhlmüller, N. Färber, M. Link, and B. Girod, "Analysis of Video Transmission over Lossy Channels", *IEEE Journal on Selected Areas in Communications*, vol. 18, no 6, pp. 1012-1032, June 2000.

[2] Z. He, and C. W. Chen, "End-to-End Quality Analysis and Modeling for Video Streaming over IP Network", *Proc. IEEE International Conference on Multimedia and Expo*, ICME 2002.

[3] L. U. Choi, W. Kellerer, and E. Steinbach, "Cross-Layer Optimization for Wireless Multi-User Video Streaming", *Proc. IEEE International Conference on Image Processing*, ICIP 2004, Singapor, Oct. 2004.

[4] M. T. Ivrlač, and J. A. Nossek, "Cross Layer Design - An Equivalence Class Approach", *Proc. IEEE International Symposium on Signals, Systems, and Electronics*, ISSSE-04, Linz, Austria, 2004.

[5] J. Ribas-Corbera, and S. Lei, "Optimal Quantizer Control in DCT Video Coding for Low Delay Video Communication", *Proc. Picture Coding Symposium*, pp. 749-754, Berlin, March 1997.

[6] J. L. H. Webb, and K. Oehler, "A Simple Rate-Distortion Model, Parameter Estimation, and Application to Real-Time Rate Control for DCT-based Coders", *Proc. International Conference on Image Processing*, vol. 2, Santa Barbara, CA, Oct. 1997, pp. 13-16.

[7] W. Ding, and B. Liu, "Rate Control of MPEG Video Coding and Recoding by Rate-Quantization modelling", *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 6, pp. 12-20, Feb. 1996.

[8] A. Gersho, "Asymptotically Optimal Block Quantization", *IEEE Trans. on Information Theroy*, vol. 25, no. 4, pp. 373-380, July 1979.

[9] S. Mallat, and F. Falzon, "Analysis of Low Bit Rate Image Transform Coding", *IEEE Trans. on Signal Processing*, vol. 46, no. 4, pp. 1027-1042, April 1998.