ACCURATE ESATIMATION OF R-D CHARACTERISTICS FOR RATE CONTROL IN REAL-TIME VIDEO ENCODING

Junfeng Bai, Chang Feng*, Qingmin Liao**, Xinggang Lin**, Xinhua Zhuang**

* Multimedia Communications and Visualization Laboratory
Department of Computer Engineering and Computer Science
University of Missouri-Columbia, MO 65211 USA

** Department of Electronic Engineering, Tsinghua University,
Beijing 100084, China

ABSTRACT

In real-time video communications, the rate control strategies must be utilized to satisfy the end-to-end delay and prevent the encoding buffer from over/underflow. In other words, to acquire the best possible video quality with a minimal quality variation in the playback video, an accurate Rate-Distortion (R-D) model of the video source is critical in optimizing the bit allocation for video coding. In the paper, an Exponential Functions based R-D model is proposed for Intra-coded frames in video coding. Numerous experiments have consistently shown that the proposed model outperforms other popular R-D models in terms of both the estimation accuracy and computation complexity, making it suitable for rate control in real-time video coding.

1. INTRODUCTION

Real-time visual communications is becoming increasingly popular with the advances in video compression and network technology. To satisfy the end-to-end delay and prevent the encoding buffer from over/underflow in real-time visual communications systems, rate control strategies are always utilized [1, 5]. In other words, to acquire the best possible video quality with a minimal quality variation in the playback video, an accurate estimation of the rate-distortion characteristics of the video source is critical in optimizing the bit allocation for video encoding. Especially, in a forward rate control system, if there were a large estimation error in one frame, it would have bad influence on the bit allocation for following frames, producing uneven video quality. In order to predict the output bit-rate and the related distortion of the encoder at the given coding parameters - quantization scales, for example - it is necessary to build an accurate source model to estimate the Rate-Distortion (R-D) characteristics of the input data. The more accurate the model, the better the bit allocation.

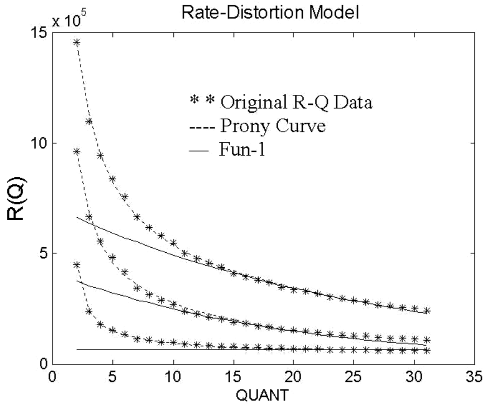
For still images or Intra-coded video frames, the R-D characteristics are difficult to acquire using closed-form models. Some traditional R-D models assume that the input data is Gaussian, Laplacian, or generalized Gaussian distributed. However, due to the non-stationary characteristics of the image signal, the accuracy of these parametric models is often too low to be useful for optimized rate control in video coding [1,3]. More advanced R-D models, such as the Quadratic model and the Cubic Interpolation model, are developed based on a priori knowledge of the Rate-Distortion characteristics of image signals, with the model parameters being derived from the actual image to be encoded [1, 2]. However, it is observed that large estimation errors may occur when the Quadratic model is used to estimate the bit-rate at small quantization scales (QUANT), which makes the model unsuitable for high bit rate video coding. Furthermore, while the Cubic model may provide more accurate estimation of R-D characteristics for the encoded image, its computational complexity in obtaining the R-D data at a reasonable number of control points (eight, for instance) is often too high to be feasible in real-time video coding.

In this paper, we propose an Exponential Functions-based R-D model for the estimation of the R-D characteristics of the images to be encoded. The proposed model is similar to the Cubic Interpolation model in that they both employ control points followed by estimation of the R-D curve through interpolation. The two differ in that our method can achieve comparable estimation accuracy using considerably less number of control points (i.e., three), which significantly reduces the computational complexity of the method and makes it suitable for real-time video coding.

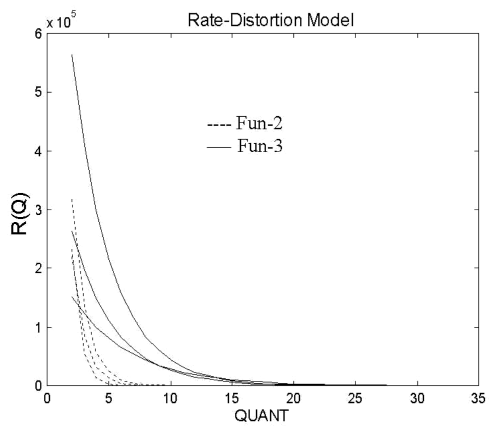
The paper is organized as follows. In section 2, the exponential functions-based R-D model for intra-coded video frames is explained in detail. In section 3, the performance of the proposed model is compared to two other models. The computational complexity of the model is analyzed in section 4 to demonstrate that the proposed method can be used for real-time video coding. Some conclusions of the paper are provided in the final section.

2. EXPONENTIAL FUNCTIONS BASED RATE-DISTORTION MODEL

In high bit-rate video coding, the quantization scales (QUANT) are changed to control the video encoding rate (R) and the corresponding encoding distortion (D) of each frame. An accurate estimation of the R-D relationship is of vital importance for achieving guaranteed end-to-end delay and smooth video playback. In this paper the relationship between R and QUANT is estimated using a novel exponential functions based model. The relationship between D and QUANT is derived via interpolation from the actual distortions of the encoded image at the QUANT values chosen for the R-QUANT estimation process.



(a)



(b)

Figure 1. Exponential functions based R-Q model

2.1 Rate- QUANT model

Some typical image Rate-QUANT (R-Q) curves are shown in Figure 1 (a). It is observed that the R-Q curves have strong non-linearity at small QUANT values (e.g., Q < 15). To analyse the characteristics of the R-Q curves, the Prony curve-fitting method is employed, which can fit any curve using a combination of N exponential functions. That is, we can build a function F(Q)

$$F(Q) = \sum_{n=1}^{N} Coef1(n) \times \exp(-Coef2(n) \times Q) \qquad (N <= 31) \tag{1}$$

such that $R(Q_i) \approx F(Q_i), \quad Q_i \in \{1,2,\ldots,31\}$ under some error measures. In our experiments, the Minimizing Absolute Maximum Error (*MinAbsMax*) principle is employed for curve fitting. The 2*N coefficients of the N exponential functions can be derived from a set of linear equations, each corresponding to the actual R-Q data at a particular QUANT value (from 1 to 31).

Through our experiments we find that three exponential curves (*Fun-1, 2, and 3* in Figure 1) are sufficient to accurately fit the original R-Q curve. To acquire the coefficients of these three exponential curves, we have to use the actual R-Q data obtained at all QUANT values (1 <= Q <= 31). Such brute-force approach, however, requires that the video frame must be encoded and decoded at every possible value of QUANT, which is not feasible in real-time processing. Instead, by observing the characteristics of the three exponential curves, we have designed a much simpler approach that dramatically reduces the computational complexity for building the R-D model with only minor performance loss.

Figure 1 shows that the first exponential function (*Fun-1*) alone fits the R-Q data nicely at large quantization scales (e.g., QUANT>10), where the R-Q curve changes slowly. The second exponential function (*Fun-2*) captures the fast-decreasing characteristics of the R-Q data at smaller quantization scales (e.g., QUANT<10). Finally, the third exponential function (*Fun-3*) can be considered as compensation for the estimation errors caused by *Fun-1* and *Fun-2*. Intuitively, it is possible to use the near-linear section (i.e., larger Q values) of the R-Q data to derive *Fun-1*, and use the non-linear section (i.e., smaller Q values) of the R-Q data to obtain *Fun-2*. *Fun-3* can then be derived based on the estimation errors generated by *Fun-1* and *Fun-2*. Finally, by adding the three functions together, we can obtain an accurate estimation of the original R-Q data at all possible quantization scales.

2.1.1 Estimation of the linear component of R-D curve

Our target function is

$$Fun1(Q) = Coef1(1) * \exp(-Coef2(1) * Q) \qquad 1 \le Q \le 31 \tag{2}$$

Since *Fun1*'s main contribution is at larger Q values (from 10 to 31), only two control points of Q1=10 and Q2=25 are used to derive the function coefficients. The set of functions used are

$$\begin{cases} Fun1(25) = R(25) \\ Fun1(10) = \alpha * R(10) \end{cases} \tag{3}$$

The adjustable factor α is used to compensate for the non-linearity of the R-Q curve near Q1. The two coefficients of *Fun-1* can then be derived using the following formulas:

$$\text{Coef2}(1) = \ln\left(Fun1\,(Q1) \,/\, Fun1\,(Q2)\right) / (Q2 - Q1) \tag{4}$$

$$\text{Coef1}(1) = \exp(0.5*(\ln(Fun1(Q1)*Fun1(Q2)) + \text{Coef2}(1)*(Q2+Q1))) \tag{5}$$

2.1.2 Estimation of the Non-linear component of R-D curve

Similarly, the coefficients *Coef1(2)* and *Coef2*(2) of *Fun2(Q)* is determined using only two control points at Q1=10 and Q2=1. The two functions used for deriving the coefficients are:

$$\begin{cases} Fun2(10) = (1-\alpha) * R(10) \\ Fun2(1) = \beta * R(1) - Fun1(1) \end{cases} \tag{6}$$

where β serves the same purpose as α. Using Equation (4,5), two coefficients *Coef1(2)* and *Coef2*(2) are derived for *Fun2(Q)* by replacing *Fun1* with *Fun2*.

2.1.3 Compensation for the estimation error

With the same principle, the third exponential function (*Fun3*), used to compensate for the estimation errors, is determined using two control points at Q1 = 1 and Q2 = 25.

$$\begin{cases} Fun3(1) = R(1) - \beta * R(1) - Fun1(1) \\ Fun3(25) = 0 \end{cases} \tag{7}$$

Finally, we acquired the estimated R-Q curve *F(Q)*

$$F(Q)=Fun1(Q)+Fun2(Q)+Fun3(Q) \quad (8).$$

2.2 Distortion- QUANT model

The MSE (Mean Square Error) is widely used as the distortion measurement in video coding. The relationship between Distortion and QUANT can be simply interpolated from the distortion values obtained by encoding the frame at the chosen QUANT values (Q = 1, 10, and 25). Through experiments, we find that the Cubic interpolation method is more accurate than Spline interpolation for all tested video sequences, as evidenced in *Figure 2*.

3. PERFORMANCE OF THE MODEL

The proposed method has been extensively experimented to compare its performance with other well-known R-D models. Test images include frames from video sequences such as “Flower Garden” (720x288 size), “Mother & Daughter” (320x288 size), and others. To test the performance of our Exponential Functions-based R-D Model, in each experiment we randomly select one frame from every sequence and encode the frame using the Intra-code mode by an MPEG2 encoder. In our experiments, the values α =95% and β =8% are empirically determined for all sequences. The model’s performance may be further improved if these two factors are adjusted adaptively according to the bit rates at the chosen control points.

First, the Quadratic Rate Distortion Model is compared [2]. Using the same original R-D data at Q={1, 10, 25}, we calculate two coefficients ***a1*** and ***b1*** for quadratic polynomial with the proposed equations in [2]. The results are shown in *Figure 3*. It is evident that our method produces significantly smaller errors for Q<15, while both methods are accurate enough for large QUANT values. Apparently, the quadratic polynomial model does not do a good job in matching the non-linearity part (i.e., small quantization scales) of the R-D curve.

Secondly, the Cubic Interpolation Model is used for performance comparison [1]. The *Relative Error* is defined as

$$Relative\_Error = \left| \frac{Estimated\_Value - Original\_Value}{Original\_Value} \right| * 100\% \quad (10)$$

Piecewise cubic polynomials are used to interpolate the bit rate related to QUANT between the chosen control points. In Cubic Interpolation, the control points of {1, 3, 5, 8, 13, 21, 31} are used, as proposed in [1]. In our model, Q={1, 10, 25} are used as control points. The results are listed in *Table 1*. The results show that our method achieves comparable prediction accuracy with the significantly reduced number of control points. Less control points means lower computational complexity of the method, which is vital for its applicability in real-time video coding. The computation complexity of our model, and the feasibility of it being used for real-time video coding, will be further discussed in the next section.

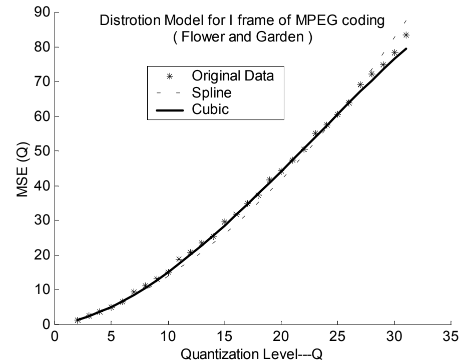


Figure 2. Estimation of the Distortion-QUANT characteristics

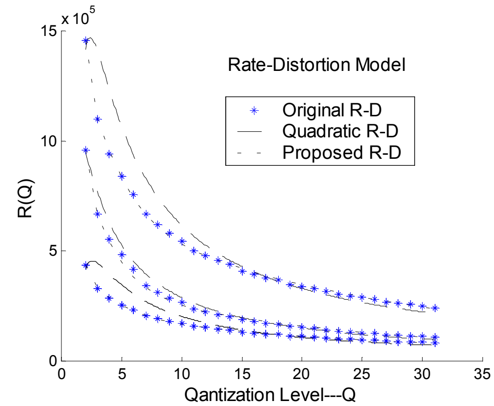


Figure 3. Comparison of the performance of our model with Quadratic Rate-Distortion Model

4. ANALYSIS OF THE MODEL’S COMPUTATION COMPLEXITY

The R-D data needed for computing the coefficients of the exponential functions for a frame is obtained via actually coding the video frame at the chosen control QUANT values. Due to the computational complexity of the method, in real-time rate control systems, the proposed R-D based rate control model is only employed in estimating the R-D characteristics for the Intra-coded frames, when the time-consuming motion estimation/compensation process is not carried out. For motion-compensated frames (i.e., Inter-frames), a simple model can be utilized to estimate the coding rate based on the assumption that the error signal is Laplacian distributed [3].

Our discussion on the feasibility of the exponential R-D model in real-time video coding is based on the assumption that the amount of computation needed for building the model is no more than that for motion estimation/compensation in Inter-coded frames. Thus, as long as motion-estimation/compensation can be achieved in real-time, our R-D model for Intra-coded frames can also be built in real-time. Generally speaking, in video coding the steps necessary for R-D modeling include scalar quantization, variable length encoding, dequantization, and inverse DCT. Here, the estimation of computation complexity for H.261 (*Table 2*) is employed to show the feasibility of our R-D model [4].

As shown in Table 2, the main computation in our R-D modelling procedure is 3*[(Quantization, Zig-zag scanning) + (Entropy coding) + (Inverse Quantization) + (Inverse DCT)] = 570 *MOPS*, whereas the computation involved in motion estimation/compensation is 608 *MOPS*. Therefore, it is clear that the proposed R-D modelling procedure for Intra-frames can be accomplished within the same time period used for motion estimation in Inter-frames. This conclusion is also valid for other standard video coding systems since the main encoding procedures are very similar to H.261.

5. CONCLUSION

In this paper, we have proposed an exponential functions based R-D model for accurate rate control in real time video coding. Our experiments have demonstrated that the model has shown high estimation accuracy with significantly lower computation complexity. The proposed model has been used for real-time rate control in MPEG VBR video encoding and transmission [5] to achieve optimized bit allocation, guaranteed end-to-end delay, and smooth video playback.

REFERENCES

[1] Liang-jin Lin, Antonio Ortega, Bit-Rate Control Using Piecewise Approximated Rate-Distortion Characteristics, IEEE Trans. Circuits and Systems for Video Technology, Vol. 8, No. 4. Aug. 1998, pp. 458-446.

[2] Tihao Chiang, Ya-Qin Zhang, A New Rate Control Scheme Using Quadratic Rate Distortion Model, IEEE Trans. Circuits and Systems for Video Technology, Vol. 7, No. 1, Feb. 1997, pp. 246-250.

[3] Hsueh-Ming Hang, Jiann-Jone Chen, Source Model for Transform Video Coder and Its Application - I: Fundamental Theory, IEEE Trans. Circuits and Systems for Video Technology, Vol. 7, No. 2, April. 1997, pp. 287-298.

[4] Vasudev Bhaskaran, Konstantinos Konstantinides, Image and Video compression standards: algorithms and architectures, Boston: Kluwer, 1997

[5] Junfeng Bai, Qingmin Liao, Xinggang Lin, Rate-Distortion models based rate control algorithm for real-time VBR video encoding, Proceeding of SPIE Vol. 4067, Visual Communications and Image processing 2000. pp. 438-446

Table 1. Comparison of the proposed Model with Cubic Interpolation Model

Data		**Exponential Function based Model**		**Cubic Interpolation Model**	
		Relative Error		***Relative Error***	
		MEAN (%)	**MAX (%)**	**MEAN (%)**	**MAX (%)**
Flower Garden		1.76	3.92	1.62	4.08
Table Tennis		3.11	8.38	3.59	8.57
Susan		3.41	23.47	4.28	10.27
Wind and Leaves		2.30	5.22	2.57	6.32
Mobile and Calendar		1.55	3.65	1.45	3.82
* Bowing		1.30	2.94	1.60	3.79
* Mother and Daughter		1.07	2.30	1.18	3.32
* Pamphlet		1.49	3.45	1.45	4.03
* Paris		1.18	2.70	0.74	2.36
Summary	***Mean***	***1.91%***	***6.23%***	***2.05%***	***5.17%***
	Var	***0.35%***	***6.71%***	***0.45%***	***2.71%***

Table 2. Estimation of the computation (*MOPS**) in H.261 video codec [4]

Compression	***MOPS***	**Decompression**	***MOPS***
Motion estimation (25 Searches in a 16*16 region)	***608***		
Entropy coding	***17***	**Entropy decoder**	***17***
2-D DCT	***60***	**Inverse DCT**	***60***
Quantization, Zig-zag Scanning	***44***	**Inverse Quantization**	***9***
Entropy coding	***17***	**Loop filtering**	***55***
Frame reconstruction	***99***		

* (*MOPS*: Million operations per second)