A NOVEL LINEAR SOURCE MODEL AND A UNIFIED RATE CONTROL ALGORITHM FOR H.263 / MPEG-2 / MPEG-4

Yong Kwan Kim, Zhihai He, and Sanjit K. Mitra

Department of Electrical and Computer Engineering
University of California, Santa Barbara, CA 93106
{zhihai, mitra}@iplab.ece.ucsb.edu

ABSTRACT

Let ρ be the percentage of zeros among the quantized transform coefficients. We discover that, in any typical video coding systems, there is always a strictly linear relationship between ρ and the actual coding bit rate R. This linearity leads to a novel and unified source model for different types of source data and different coding systems, such as H.263 [1], MPEG-2 [2], and MPEG-4 [3]. The proposed linear source model is much simpler, but much more accurate than other source models reported in the literature. Based on this source model, a unified rate control algorithm is proposed for the above three video coding systems. Despite its extreme simplicity, the proposed algorithm outperforms other rate control algorithms by providing more accurate and robust rate control.

1. INTRODUCTION

In visual communication over narrow-band or time-varying channels, rate control is very important to ensure the successful transmission of the coded video data through the communication channel. The key issue in rate control is to model and estimate the rate - quantization (R-Q) behavior of the video encoder [4].

The classical rate-distortion (R-D) formula [5] does not work for transform coding at low bit rates [6]. Several source models and rate control algorithms have been proposed in the literature [4, 6, 7, 8]. Most of them [4, 6, 8] are the modified versions of the classical R-D formula. These algorithms either have very high computational complexity [7], or suffer from large estimation or control error due to their inaccurate source models, especially at scene changes [4, 6, 8]. In addition, they only deal with specific coding algorithms or specific types of source data [4, 6]. In this work, our objective is to develop a low-complexity, accurate and unified rate control algorithms for all typical video coding systems, such as H.263 [1], MPEG-2 [2] and MPEG-4 [3].

To our best knowledge, all the source models reported in the literature [4, 6, 7, 8] try to find the best expression for the coding bit rate R in terms of of the quantization stepsize q. That is to say, the rate functions are defined and modeled in the q-domain. It is well known that zeros play a key role in transform coding of images and videos. Let ρ be the percentage of zeros among the quantized transform coefficients. Obviously, ρ monotonically increases with q, which implies there is a one-to-one mapping between them. Hence, mathematically, the coding bit rate R is also a function of ρ, denoted by $R(\rho)$. In this work we propose to study the rate function in the ρ-domain instead of the q-domain.

In the ρ-domain, we discover that the rate function $R(\rho)$ has an extremely simple form. To be more specific, for any typical transform coding systems such as H.263, MPEG-2 and MPEG-4, $R(\rho)$ is always a linear function [1]. This linearity leads to a novel and unified source model in the ρ-domain for different coding systems. It is much simpler, but much more accurate than other source models proposed in the literature [4, 6, 7, 8]. Based on this linear source model, a unified rate control algorithm is proposed for H.263, MPEG-2 and MPEG-4 video coding. Our simulation results show that the proposed algorithm outperforms the TMN8 [4] and VM7 [8] rate control algorithms by providing much more accurate and robust rate control.

2. A UNIFIED LINEAR SOURCE MODEL

In this section, we show that, in any typical video coding systems, for any type of video frames, such as I-,

This work was supported by a University of California MICRO grant with matching support from Lucent Technologies, National Semiconductor, Tektronix Corporation, and Xerox Corporation.

[1] Our simulations show that this linearity also holds for wavelet-based image / video coding systems.

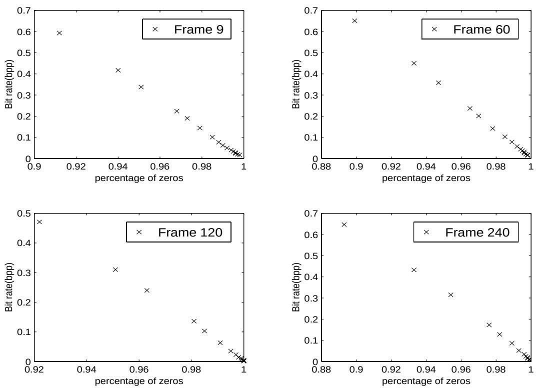

Figure 1: The linear relationship between the coding bit rate and the percentage of zeros. The test frames are from the Foreman QCIF video.

P- and B- frames, there is always a strictly linear relationship between R and ρ. To this end, a series of experiments have been performed as described in the following.

With the H.263 video codec, we encode the test video sequence at a series of quantization stepsizes. In each video frame, for any given quantization stepsize q, let R be the coding bit rate excluding the motion vectors (MV) and the header information bits. It should be noted that the amount of bits for MV and header information is already determined before rate control and quantization. We can not change it during the rate control process. In Fig. 1, we plot $R(\rho)$ for several frames from the Foreman video sequence. It can be seen that there is a very strong linear relationship between R and ρ. To be more specific, R is proportional to $1-\rho$.

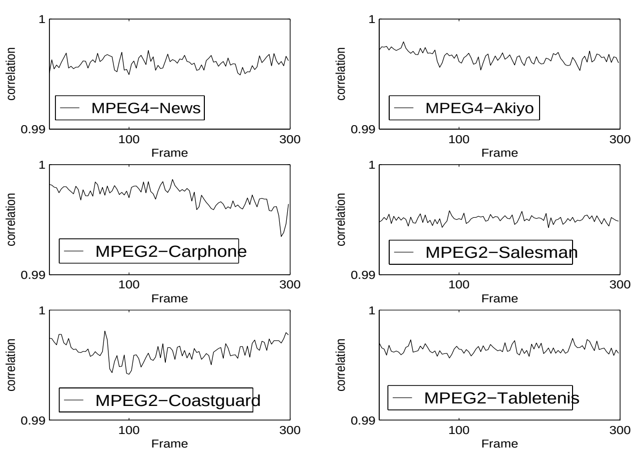

Figure 2: The correlation coefficient (inverse) of each frame between the coding bit rate R_c and ρ.

To justify this linear relationship more systematically, we study the correlation coefficient between R and ρ, denoted by $\mathcal{C}(\rho, R)$. In Fig. 2, we plot the values of $-\mathcal{C}(\rho, R)$ for each frame of Akiyo and News coded by MPEG-4, Carphone, Salesman, Coastguard and Tabeltennis coded by MPEG-2. It should be noted that in MPEG-4, R does not include the bits for shape information. From Fig. 2 it can be seen that $-\mathcal{C}(\rho, R)$ is always larger than 0.99. For most of the frames, it is larger than even 0.995 which is extremely close to 1. This implies there is a strictly linear relationship between R and ρ. Note that when ρ is 1.0, which means all the coefficients are quantized to zeros, the coding bit rate R should also approach zero. Therefore, we have the following linear source model

$$R(\rho) = \theta \cdot (1 - \rho). \tag{1}$$

where θ is a constant for each frame. Our extensive simulations over a variety of video sequences show that the above linear source model always holds.

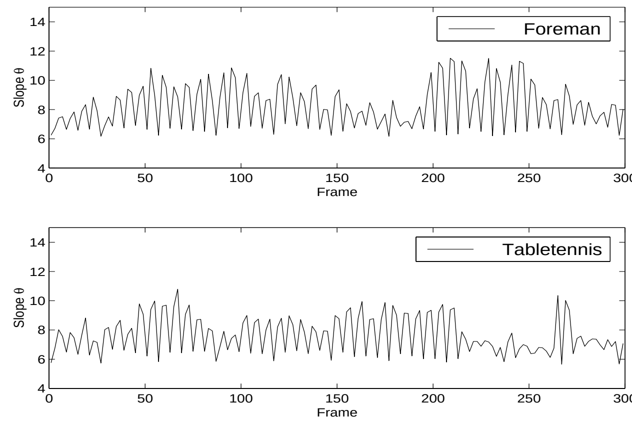

Figure 3: The slope θ of each frame in Foreman coded by MPEG-2.

3. THE MAPPING BETWEEN Q AND ρ

The coding bit rate R is controlled by adjusting the quantization stepsize q of the encoder. As discussed in Section 1, there is a one-to-one mapping between q and ρ. This mapping can be easily computed from the distribution of the DCT coefficients. Let us consider the H.263 encoder as an example. Let Δ be the dead-zone threshold of the quantizer. In H.263, Δ is q for intra-macroblocks and $1.25q$ for inter-macroblocks. Let $\mathcal{D}_0(x)$ and $\mathcal{D}_1(x)$ be the distributions of the DCT coefficients in the intra- and inter- macroblocks, respectively. For any q, the corresponding ρ can be computed as follows,

$$\rho(q) = \frac{1}{M} \sum_{|x| \le q} \mathcal{D}_0(x) + \frac{1}{M} \sum_{|x| \le 1.25q} \mathcal{D}_1(x). \tag{2}$$

where M is number of coefficients in the current frames. The above computation only involves several additions. In the MPEG-2 quantization scheme, a perceptual quantization matrix, denoted by $[W_{ij}]$, is employed. In this case, we first scale each DCT coefficient by its perceptual weight W_{ij} and then treat the quantization as uniform. In this way, a similar formula as Eq. (2) can be developed based on the distribution of the DCT coefficients. In our rate control algorithm, the mapping between q and ρ is stored as a look-up table.

4. ESTIMATION OF θ

The only parameter in our linear source model given in Eq. (1) is the slope θ. In Fig. 3 we plot the slope θ for each frame of Foreman and Tabletennis coded by MPEG-2. It can be seen that θ has a large variation. In this work, we propose an adaptive estimation algorithm to estimate the value of θ for each frame. Let N_m be the number of the coded macroblocks in the current frame. Note that in a 16×16 macroblock, there are totally 384 luminance and chrominance coefficients. Let R_m be the number of bits used to encode these N_m macroblocks. We denote the number of zeros in these macroblocks by ρ_m. Based on Eq. (1), θ can be estimated as follows,

$$\theta = \frac{R_m}{384 \cdot N_m - \rho_m}. \tag{3}$$

The estimated θ is then applied to the current macroblock.

5. RATE CONTROL ALGORITHM

With the linear source model and the adaptive estimation of θ, the rate control becomes very simple and straightforward. Let the target bit rate (in bits) per frame be R_T. Let the encoder buffer size be B_T and the number of bits in the buffer be B_0. The available bits for coding the current frame is

$$R = R_T - B_0 + \alpha \cdot B_T, \tag{4}$$

where the target buffer level α is by default set to be 0.2. Let M be the number of macroblocks in one video frame. For QCIF videos, M is 99. The quantization parameter is determined in the following steps:

Step 1: Initialization. Before encoding the first macroblock, set $N_m = R_m = \rho_m = 0$. Compute the distributions $\mathcal{D}_0(\mathrm{x})$ and $\mathcal{D}_1(\mathrm{x})$. Set $\theta = 7$ which is its average value.

Step 2: Determine the quantization parameter q. According Eq. (1), the number of zeros to be produced by quantizing the rest macroblocks should be

$$\rho = 384 \cdot (M - N_m) - \frac{R - R_m}{\theta}. \tag{5}$$

Based on the one-to-one mapping between ρ and q, the stepsize q is then determined.

Step 3: Update. Let ρ_0 and R_0 be the number of zeros and number of bits produced by the current macroblock. Set $\rho_m = \rho_m + \rho_0$, $R_m = R_m + R_0$, and $N_m = N_m + 1$. If $N_m \geq 10$, update the value of θ according to Eq. (3). At the same time, subtract the frequencies of the DCT coefficients in the current macroblock from $\mathcal{D}_0(\mathrm{x})$ or $\mathcal{D}_1(\mathrm{x})$.

6. EXPERIMENTAL RESULTS

The proposed rate control algorithm has been implemented in the H.263, MPEG-2 and MPEG-4 video coders. In the first experiment, the test videos are Foreman, Salesman, News and Tabletennis. The channel rates are 48 kbps, 32 kbps, 24 kbps and 48 kbps, respectively. The frame rate is 10 fps. In Fig. 4, we plot the number of bits in the encoder buffer when the proposed algorithm (solid line) and the TMN8 rate control algorithm (dotted line) are applied. It can be seen that the proposed rate control algorithm meets the target bit rate more accurately and maintains a much steadier buffer level than TMN8. It should be noted that from [4] we know TMN8 significantly outperforms VM7 [8]. In Table 1, we list the average PSNR of the luminance component. The improved PSNR performance of the proposed algorithm is due to its more accurate and robust rate control and less skipped frames. In Fig. 5, we plot the number of bits in the encoder buffer when the proposed rate control algorithm is applied to the MPEG-4 codec. The corresponding simulation results of TMN8 can be found in [4]. It can be seen that the buffer level is almost constant throughout the coding process, which is highly desirable in low-delay real-time visual communications. To test the rate control performance of the proposed algorithm in MPEG-2 video coding, as in [7], we set the target bit rates of I, P and B frames to be 22000, 11000 and 5000 bits, respectively. In Fig. 6, we plot the bits produced by each frame. It can be seen that the proposed rate control algorithm matches the target bit rate very well. The relative error is less than 2%.

7. CONCLUDING REMARKS

There are two major contributions in this work. First, we have proposed a novel framework for source modeling by studying the rate function in the ρ-domain instead of the traditional q-domain. A unified linear source model have been developed for typical transform video coding systems. The proposed source model is very simple but very accurate. Second, a unified rate

Table 1: Comparison of the PSNR performance between TMN8 and the proposed rate control algorithm.

Video name	Target Rate	PSNR TMN8	PSNR This work
Foreman	48 Kbps	30.31	30.41
Salesman	32 Kbps	30.92	30.71
News	24 Kbps	30.28	30.73
Tabletennis	48 Kbps	29.43	29.55

control algorithm is proposed for different video coding systems. Despite its simplicity, the proposed algorithm outperforms other rate control algorithms by providing more accurate and robust rate control.

8. REFERENCES

[1] Telenor codec, "ITU-T/SG-15, video codec test model, TMN5," Telenor Research, January 1995.

[2] D. LeGall, "MPEG: A video compression standard for multimedia application," *Commun. ACM*, vol. 34, pp. 46–58, April 1991.

[3] T. Sikora, "The MPEG-4 video standard verification model," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 7, pp. 19 – 31, February 1997.

[4] J. Ribas-Corbera and S. Lei, "Rate control in DCT video coding for low-delay communications," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 9, pp. 172 - 185, February 1999.

[5] H. Gish and J. N. Pierce, "Asymptotically efficient quantizing," *IEEE. Trans. on Inform. Theory*, vol. IT-14, pp. 676-683, Sept. 1968.

[6] H. -M. Hang and J. -J. Chen, "Source model for transform video coder and its application – Part I: Fundamental theory," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 7, pp. 287 – 298, April 1997.

[7] W. Ding and B. Liu, "Rate control of MPEG video coding and recording by rate-quantization modeling," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 6, pp. 12–20, Feb. 1996.

[8] T. Chiang, Y. -Q. Zhang, "A new rate control scheme using quadratic rate distortion model," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.7, pp. 246 – 250, Feb. 1997.

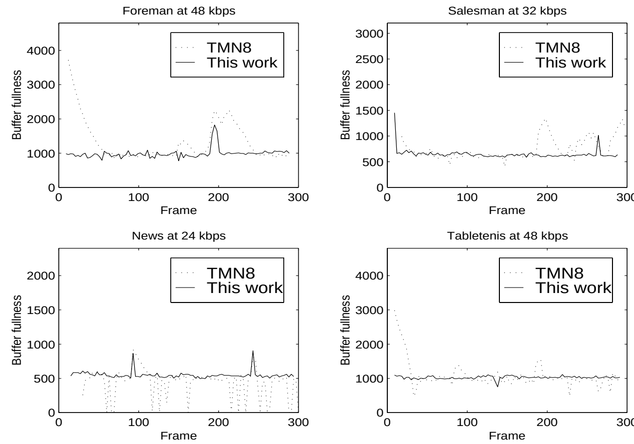

Figure 4: The number of bits in the buffer when the proposed algorithm (solid line) and the TMN8 rate control algorithm (dotted line) are appplied to H.263.

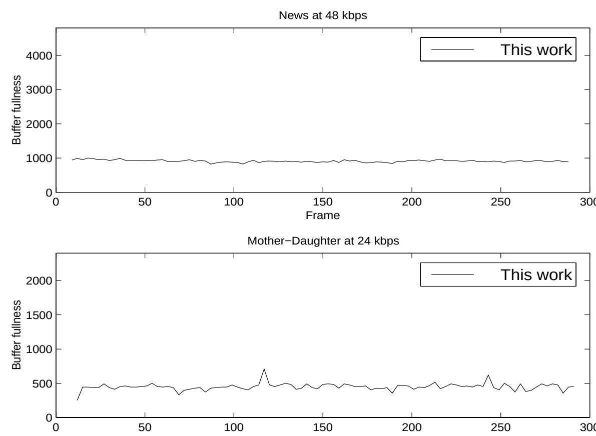

Figure 5: The number of bits produced by each encoded frame (top) when the proposed algorithm is applied to the MPEG-4 codec.

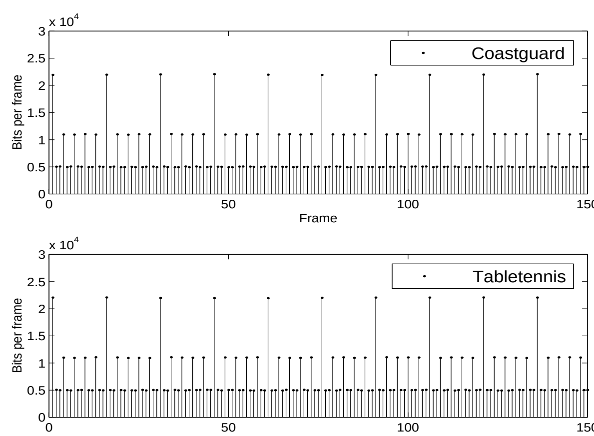

Figure 6: The number of bits produced by each encoded frame (top) when the proposed algorithm is applied to the MPEG-2 codec.