Low-Delay Rate Control for DCT Video Coding via ρ-Domain Source Modeling

Zhihai He, Member, IEEE, Yong Kwan Kim, Member, IEEE, and Sanjit K. Mitra, Life Fellow, IEEE

Abstract—By introducing the new concepts of characteristic rate curves and rate-curve decomposition, a generic source-modeling framework is developed for transform coding of videos. Based on this framework, the rate-quantization (R-Q) and distortion-quantization (D-Q) functions (collectively called R-D functions in this work) of the video encoder can be accurately estimated with very low computational complexity before quantization and coding. With the accurate estimation of the R-Q function, a frame-level rate control algorithm is proposed for DCT video coding. The proposed algorithm outperforms the TMN8 rate control algorithm by providing more accurate and robust rate regulation and better picture quality. Based on the estimated R-D functions, an encoder-based rate-shape-smoothing algorithm is proposed. With this smoothing algorithm, the output bit stream of the encoder has both a smoothed rate shape and a consistent picture quality, which are highly desirable in practical video coding and transmission.

Index Terms—Rate control, rate-distortion estimation, rate smoothing, source modeling, video coding.

I. INTRODUCTION

VIDEO CODING and transmission has become a key and integrated function block in multimedia communication over the Internet. The output bit rate of the video encoder is controlled by its quantization parameters. The core problem in rate control is to estimate the R-D functions of the video encoder [1], [2]. Based on the estimated R-D functions, we can select the corresponding quantization parameter of the encoder to achieve the target bit rate or picture quality. Several R-Q models for DCT-based video encoders have been proposed in the literature [3]–[8]. Some of them [5], [7] are based on the classical entropy formula [14], [15]. Others are based on different mathematical functions, such as power [3] and polynomial [6]. The R-D model proposed in [8] employs a combination of the logarithmic and the quadratic functions. Usually, these models have several parameters. The model parameters are empirically estimated from the previous coding results of the coding system and the statistical properties of the input source data. These models and the corresponding rate control algorithms either have very high computational complexity [3], [4], or suffer from performance degradation at scene changes, especially in active videos at low bit rates [5], [6], [8].

A. ρ-Domain R-D Analysis

All of the R-D models reported in the literature [3]–[5], [7], [8] attempt to find the formula relating the coding bit rate $R$ and distortion $D$ with the quantization parameter $q$. A study of the rate and distortion as functions of $q$ is termed $q$-domain R-D analysis. In other words, the conventional R-D analysis is carried out in the $q$ domain.

In transform coding of images and videos, we observe that the percentage of the zeros among the quantized transform coefficients, denoted by $\rho$, has a critical effect on the coding bit rate $R$, especially at low bit rates. Note that, if we assume the distribution of the transform coefficients is continuous and positive, $\rho$ monotonically increases with the quantization parameter $q$. This implies that there is a one-to-one correspondence between $q$ and $\rho$. Hence, mathematically, $R$ and $D$ are also functions of $\rho$, denoted by $R(\rho)$ and $D(\rho)$, respectively. A study of the $R$ and $D$ as functions of $\rho$ is termed $\rho$-domain R-D analysis. In this paper, based on the $\rho$-domain analysis methodology, we develop a generic R-D modeling and control framework for transform coding of videos.

B. Proposed Source-Modeling Framework

It is well known that the performance of a coding system is determined by both the characteristics of the input source data and the coding algorithm. The proposed source model is developed by two major steps. In the first step, we define two rate curves (called characteristic rate curves) in the $\rho$ domain to characterize the transform coefficients. We find that these two rate curves have interesting properties in the $\rho$ domain. Based on these properties, we show that these two rate curves can be directly estimated from the distribution of the transform coefficients. In the second step, we introduce the new concept of rate-curve decomposition to model the coding algorithm. The actual rate curve in the $\rho$ domain is represented by a linear combination of the two characteristic rate curves. Based on these two steps, a unified source model is developed for different video coding systems, such as H.263 [1] and MPEG-2 [2].

The proposed source model has very low computational complexity. Using this model, we can estimate the R-D functions accurately before quantization and coding, and control the coding
bit rate and the picture quality robustly according to the network conditions and application requirements. Our extensive simulation results show that the proposed rate control algorithm outperforms the TMN8 by providing more accurate rate regulation and better picture quality. Based on the estimated R-D functions, an encoder-based rate-shape smoothing algorithm is proposed to make a tradeoff between the variable-bit-rate (VBR) coding and the constant-bit-rate (CBR) coding. Using the proposed rate smoothing algorithm, we can control the video encoder such that the output bit stream has both a smoothed rate shape and a consistent picture quality.

C. Paper Organization

The paper is organized as follows. In Section II, we define the characteristic rate curves, discuss their properties, and propose a fast algorithm to estimate them. In Section III, we introduce the new concept of rate-curve decomposition. In Section IV, based on the proposed source-modeling framework, a rate control algorithm for DCT video coding is developed. In Section V, an encoder-based rate-shape-smoothing algorithm is presented. In Section VI, we implement the rate control and smoothing algorithm in the H.263 video encoder and compare its performance with the well-known TMN8 [5] rate control algorithm. Concluding remarks are given in Section VII.

II. CHARACTERISTIC RATE CURVES

A. Definition

Standard video coding systems, such as H.263 and MPEG-2, are based on motion compensation (MC) and DCT. The MC difference picture is generated after motion estimation and compensation. DCT is then applied to the difference picture. The DCT coefficients in each 8 × 8 block are then quantized and coded by the video coding algorithm. We can see that, for a given coding algorithm, the R-D functions are totally determined by the characteristics of the DCT coefficients. The definitions of the characteristic rate curves are based on the following two observations: First, we observe that the zeros play a key role in transform coding of images and videos, especially at low coding bit rates. Therefore, it is necessary and beneficial to model these zero coefficients separately. In other words, we need to classify the DCT coefficients into two groups: zero and nonzero coefficients, and model them separately. For this reason, we introduce two characteristic rate curves, denoted by \( Q_z(\rho) \) and \( Q_{\text{uw}}(\rho) \), to characterize the zero and nonzero coefficients, respectively. Second, we believe that in most typical transform coding systems [1], [13], [16] the coding of the nonzero coefficients is comparable to binary representation, and the coding of zeros is comparable to the binary representation of their run length. For example, in the wavelet-based SPIHT coding algorithm [13], the coefficients are coded bit-plane by bit-plane [13]. In MPEG [2] and H.263 [1] coding, each nonzero coefficient is jointly Huffman coded with the run length of its preceding zeros. Note that the length of Huffman codeword increases with the both the size of the nonzero coefficient and the size of the zeros run length. Therefore, we believe that the binary representation of the nonzero coefficients and the zeros run length numbers will give us the most valuable information about the coding behavior of the DCT coefficients.

Based on the above two observations, we define the characteristic rate curves using the following “pseudocoding” scheme. After MC and DCT, we quantize the DCT coefficients with a quantization parameter \( q \). The quantized coefficients are rearranged into a 1-D array \( L \) in a zig-zag scan order inside each macroblock and in a blockwise raster scan order at the block level. For each consecutive string of zeros between two nonzero coefficients in \( L \), we count their run length. We then sum up the sizes of all run length numbers, and denote the sum by \( Q_z \). Here, the size of a nonzero integer is the number of bits for its sign-magnitude representation. For example, the size of +17 is 6, which includes one sign bit and five bits for the magnitude. Likewise, let \( Q_{\text{uw}} \) be the sum of the sizes of all nonzero coefficients. Suppose the frame size is \( M \). Then

\[
Q_{\text{uw}} = \frac{1}{M} Q_{\text{uw}}, \quad Q_z = \frac{1}{M} Q_z
\]

can be regarded as the “pseudocoding” bit rates for the zero and nonzero coefficients, respectively. Obviously, both \( Q_{\text{uw}} \) and \( Q_z \) are functions of the quantization parameter \( q \), denoted by \( Q_{\text{uw}}(q) \) and \( Q_z(q) \). Let \( \rho \) be the percentage of the zeros among the quantized DCT coefficients. We know that \( \rho \) monotonically increases with \( q \). Therefore, mathematically, \( Q_{\text{uw}} \) and \( Q_z \) are also functions of \( \rho \), denoted by \( Q_{\text{uw}}(\rho) \) and \( Q_z(\rho) \), respectively. These two functions are termed the characteristic rate curves of the DCT coefficients.

B. Statistical Properties

To study the statistical properties of \( Q_{\text{uw}}(\rho) \) and \( Q_z(\rho) \), we generate these two rate curves according to their definitions and plot them for various pictures. In the following, let us consider two video sequences “Foreman” and “Salesman” as examples. We run the H.263 coder [18] on these two videos at a fixed quantization parameter 8. For each video, we output 30 sample MC difference pictures. Each sample picture is taken at every ninth frame. Note that the first one is an I-picture without motion compensation. We plot \( Q_{\text{uw}}(\rho) \) and \( Q_z(\rho) \) for each sample picture from “Foreman” and “Salesman” in Figs. 1 and 2, respectively.

Two observations can be made from these plots. First, although the sample pictures are quite different from each other, their characteristic rate curves have almost the same pattern. It is well known that the I- and P-pictures have different statistics. But in the \( \rho \) domain, their characteristic rate curves share almost the same pattern. For comparison, we also plot the two rate curves in the \( q \) domain: \( Q_{\text{uw}}(q) \) and \( Q_z(q) \) in Figs. 3 and 4, respectively. It can be seen that, in the \( q \) domain, these two rate curves vary from picture to picture. For example, the plots “fm-28” and “fm-190” in Fig. 3, and the plots “sm-82” and “sm-118” in Fig. 4 are quite different from each other. In addition, the plots for the I-pictures are quite different from those for the P-pictures. In active videos, this type of variation becomes even larger. As a result, it is difficult to develop an accurate source model in the \( q \) domain because the source model has to deal with the large picture-dependent variation. However, in the \( \rho \) domain, this type of variation is significantly smaller.
Fig. 1. Plots of $Q_{\text{ref}}(\rho)$ (solid line) and $Q_2(\rho)$ (dash-dot line) for the 30 sample difference pictures from “Foreman.” The $x$ axis represents the percentage of zeros $\rho$ while the $y$ axis represents the bit rate. All the subplots have the same coordinate system as the one at the bottom-left corner.

Fig. 2. Plots of $Q_{\text{ref}}(\rho)$ (solid line) and $Q_2(\rho)$ (dash-dot line) for the 30 sample difference pictures from “Salesman.” The $x$ axis represents the percentage of zeros $\rho$ while the $y$ axis represents the bit rate. All the subplots have the same coordinate system as the one at the bottom-left corner.

One conclusion can also be drawn from the above comparison: the coding bit rate of a picture is much more closely related to the percentage of zeros $\rho$ than to the quantization parameter $q$. For this reason, we propose to study the coding behavior in the $\rho$ domain instead of the $q$ domain.

Our second observation is that, for each sample picture, $Q_{\text{ref}}(\rho)$ is almost a straight line. Note that when the percentage of zeros is 100%, $Q_{\text{ref}}$ becomes zero according to its definition. This implies that the straight line must pass through the point [1.0, 0.0]. Therefore, we have the following expression:

$$Q_{\text{ref}}(\rho) = \theta \cdot (1 - \rho)$$

where $\theta$ is a constant. In our extensive simulation over various video sequences, the above two observations have been found to hold.

C. Estimation of $Q_{\text{uz}}(\rho)$

Since $Q_{\text{ref}}(\rho)$ is modeled as a straight line passing through [1.0, 0.0], we only need to compute another point on it to estimate the whole rate function. In the following, we take the H.263 coding system as an example to explain the estimation process of $Q_{\text{uz}}(\rho)$.

The H.263 quantizer is essentially a uniform threshold quantizer (UTQ) [1]. Let $\Delta$ be the UTQ dead zone threshold. (In general, $\Delta$ is proportional to $q$.) For any DCT coefficient $x$, its UTQ output index is given by

$$\text{UTQ}[x; \Delta] = \begin{cases} 0, & \text{if } |x| \leq \Delta \\ \left\lfloor \frac{x + \Delta}{q} \right\rfloor, & \text{if } x > +\Delta \\ \left\lfloor \frac{x - \Delta}{q} \right\rfloor, & \text{if } x < -\Delta. \end{cases}$$

In H.263 video coding, there are two types of macroblocks (MBs): intra-coded and inter-coded. They have slightly
different quantization schemes. To be more specific, the quantization index of $x$ in the H.263 quantization scheme is given by (4), shown at the bottom of the page. Note that the range of the unquantized dc coefficient is 0–2040, which implies that the range of its differential value is 2040 to 2040. In H.263 coding, the dc coefficients from intra-coded MBs are quantized by a uniform quantizer with fixed step size 8 and coded with a fixed number of bits. Therefore, we need not consider these dc coefficients when defining the characteristic rate curves. In other words, when we scan the picture to form the 1-D array $\mathcal{L}$, these dc coefficients are skipped. Their coding bits will be automatically compensated by our rate-curve decomposition scheme addressed in the next section.

In the current video frame, let the distribution of the DCT coefficients from the intra- and inter-coded macroblocks be $D_0(x)$ and $D_1(x)$, respectively. In the reference H.263 video codec [18], the DCT coefficients are rounded to integers. Therefore, $D_0(x)$ and $D_1(x)$ are actually histograms of the DCT coefficients. According to the definition of $Q_{\text{nz}}$, each nonzero quantized DCT coefficient is independently “coded” by its binary representation. Therefore, for any given quantization parameter $q_0$, the corresponding $Q_{\text{nz}}$ can be directly computed from $D_0(x)$ and $D_1(x)$ as follows:

$$Q_{\text{nz}}(q_0) = \sum_{|x| \geq 2q_0} S(\text{UTQ}[2q_0, 2q_0; x]) \cdot D_0(x) + \sum_{|x| < 2q_0} S(\text{UTQ}[2q_0, 2.5q_0; x]) \cdot D_1(x)$$

(5)

where $S(\cdot)$ represents the size of an integer. Meanwhile, the corresponding percentage of zeros $\rho$ is computed as follows:

$$\rho(q_0) = \frac{1}{M} \sum_{|x| < 2q_0} D_0(x) + \frac{1}{M} \sum_{|x| < 2.5q_0} D_1(x).$$

(6)

The slope $\theta$ of $Q_{\text{nz}}(\rho)$ is obtained by

$$\theta = \frac{Q_{\text{nz}}(q_0)}{1 - \rho(q_0)}.$$  

(7)

The whole rate curve $Q_{\text{nz}}(\rho)$ is then constructed by the linear model in (2).

D. Correlation Model

In the following, we show that there is a strong correlation between $Q_{\text{nz}}(\rho)$ and $Q_{\text{z}}(\rho)$. To study the correlation between two curves, we first define feature variables for each curve, then study the correlation between these feature variables. The feature variable for $Q_{\text{nz}}(\rho)$ is its slope $\theta$. The feature variables of $Q_{\text{z}}(\rho)$ are its function values at $\rho = 0.89, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97$, and 0.98. Let us consider the characteristic rate curves plotted in Figs. 1 and 2. For each $p_i, 1 \leq i \leq 9$, there are 60 pairs of $[\theta, Q_{\text{z}}(p_i)]$ for the 60 sample MC difference pictures, which are depicted in Fig. 5. Let us denote the correlation coefficient between $\theta$ and $Q_{\text{z}}(p_i)$ by $C_\theta(\theta, Q_{\text{z}}(p_i))$. The values of $C_\theta(\theta, Q_{\text{z}}(p_i))$ are listed in Table I. We can see that there is a strong correlation between $\theta$ and $Q_{\text{z}}(p_i)$. In our extensive simulations over various video sequences, the above correlation has been found to exist. Obviously, this strong correlation enables us to estimate $Q_{\text{z}}(p_i)$ accurately. For the sake of higher estimation accuracy, we use the following cubic model:

$$Q_{\text{z}}(p_i) = A_0 \theta^3 + B_0 \theta^2 + C_\theta \theta + D_i$$

(8)

to estimate the values of $Q_{\text{z}}(p_i)$. The coefficients $A_i$, $B_i$, $C_i$, and $D_i$ are obtained by statistical regression and listed in Table II. The corresponding correlation model given by (8) is also depicted in Fig. 5. Note that these model coefficients are fixed in our rate-estimation algorithm once they are obtained from the sample training pictures.

III. RATE-CURVE DECOMPOSITION

A. Basic Idea

In digital signal processing (such as Fourier transform), we often represent a signal to be studied by a weighted sum of the basis signals which have well-known properties. This method is referred as signal decomposition and analysis. In this paper, we

TABLE I

<table>
<thead>
<tr>
<th>(p_i)</th>
<th>0.89</th>
<th>0.91</th>
<th>0.92</th>
<th>0.93</th>
<th>0.94</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C(\theta, Q_{\text{nz}}(p_i))$</td>
<td>-0.88</td>
<td>-0.92</td>
<td>-0.93</td>
<td>-0.95</td>
<td>-0.95</td>
</tr>
<tr>
<td>$\rho(p_i)$</td>
<td>0.95</td>
<td>0.96</td>
<td>0.97</td>
<td>0.98</td>
<td></td>
</tr>
</tbody>
</table>

TABLE II

<table>
<thead>
<tr>
<th>(p_i)</th>
<th>$A_i$</th>
<th>$B_i$</th>
<th>$C_i$</th>
<th>$D_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.89</td>
<td>-0.070</td>
<td>+0.953</td>
<td>-4.437</td>
<td>+7.236</td>
</tr>
<tr>
<td>0.91</td>
<td>-0.013</td>
<td>+0.210</td>
<td>-1.183</td>
<td>+2.475</td>
</tr>
<tr>
<td>0.92</td>
<td>+0.011</td>
<td>-0.115</td>
<td>+0.244</td>
<td>+0.392</td>
</tr>
<tr>
<td>0.93</td>
<td>+0.013</td>
<td>-0.143</td>
<td>+0.332</td>
<td>+0.289</td>
</tr>
<tr>
<td>0.94</td>
<td>+0.036</td>
<td>-0.443</td>
<td>+1.638</td>
<td>-1.622</td>
</tr>
<tr>
<td>0.95</td>
<td>+0.030</td>
<td>-0.340</td>
<td>+1.129</td>
<td>-0.828</td>
</tr>
<tr>
<td>0.96</td>
<td>-0.011</td>
<td>+0.239</td>
<td>-1.578</td>
<td>+3.330</td>
</tr>
<tr>
<td>0.97</td>
<td>-0.109</td>
<td>+1.519</td>
<td>-7.131</td>
<td>+11.310</td>
</tr>
<tr>
<td>0.98</td>
<td>-0.026</td>
<td>+0.451</td>
<td>-2.518</td>
<td>+4.613</td>
</tr>
</tbody>
</table>

\[
I[x] = \begin{cases} 
\text{Round} \left( \frac{x}{8} \right), & \text{if } x \text{ is a dc coefficient in an intra-coded MB} \\
\text{UTQ}(2q, 2q; x), & \text{if } x \text{ is an AC coefficient in an intra-coded MB} \\
\text{UTQ}(2q, 2.5q; x), & \text{if } x \text{ is a coefficient in an inter-coded MB},
\end{cases}
\]
Fig. 5. Illustration of the strong correlation between $Q_{2}(p_i)$ and $\theta$. For each $p_i$, we plot the 60 pairs of [$\theta$, $Q_{2}(p_i)$] for the 60 sample pictures. Each subplot is titled by the respective value of $p_i$.

apply this decomposition methodology to analyze and estimate the R-Q behavior of a video encoder.

In the previous section, we have defined two characteristic rate curves, $Q_{cw}(\rho)$ and $Q_{c}(\rho)$, for the DCT coefficients. Our experimental results and statistical analysis show that they have unique properties in the $\rho$ domain. These two rate curves characterize the R-Q behavior of the input source data which determines the performance of the coding system. In our rate-curve decomposition scheme, the rate function of the coding system is represented by a linear combination of $Q_{cw}(\rho)$ and $Q_{c}(\rho)$, which serve as the basis functions.

B. Decomposition Scheme

Let $R_{H}(\rho)$ be the rate function of the H.263 encoder in the $\rho$ domain. As discussed above, we approximate $R_{H}(\rho)$ by $\hat{R}_{H}(\rho)$, which is a linear combination of $Q_{cw}(\rho)$ and $Q_{c}(\rho)$

$$\hat{R}_{H}(\rho) = \xi_{1}(\rho) \cdot Q_{cw}(\rho) + \xi_{2}(\rho) \cdot Q_{c}(\rho) + \xi_{3}(\rho)$$  \hspace{1cm} (9)

where $\xi_{1}(\rho)$, $\xi_{2}(\rho)$, and $\xi_{3}(\rho)$ are selected such that $\hat{R}_{H}(\rho)$ is as close to $R_{H}(\rho)$ as possible. In practice, we only need to make sure that at some control points, e.g., $\{\rho_i | 1 \leq i \leq 9\}$, these two curves are very close to each other. In other words, we only need to determine $\xi_{1}(\rho_i)$, $\xi_{2}(\rho_i)$, and $\xi_{3}(\rho_i)$ such that the error between $\hat{R}_{H}(\rho_i)$ and $R_{H}(\rho_i)$ is minimized. To determine the value of $\xi_{1}(\rho_i)$, $\xi_{2}(\rho_i)$, and $\xi_{3}(\rho_i)$, we need to solve the following linear regression equation:

$$R_{H}(\rho_i) = \xi_{1}(\rho_i) \cdot Q_{cw}(\rho_i) + \xi_{2}(\rho_i) \cdot Q_{c}(\rho_i) + \xi_{3}(\rho_i), \hspace{1cm} 1 \leq i \leq 9.$$  \hspace{1cm} (10)

Note that, for each sample difference picture, $Q_{cw}(\rho_i)$ and $Q_{c}(\rho_i)$ can be estimated by the algorithm developed in Section II, and $R_{H}(\rho_i)$ can be generated by running the H.263 coding algorithm over this picture. The values of $\xi_{1}(\rho_i)$, $\xi_{2}(\rho_i)$, and $\xi_{3}(\rho_i)$ are obtained by linear regression and listed in Table III.

<table>
<thead>
<tr>
<th>$\rho_i$</th>
<th>$\xi_{1}(\rho_i)$</th>
<th>$\xi_{2}(\rho_i)$</th>
<th>$\xi_{3}(\rho_i)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.89</td>
<td>1.2151</td>
<td>-0.4438</td>
<td>0.9005</td>
</tr>
<tr>
<td>0.91</td>
<td>0.8089</td>
<td>-0.5030</td>
<td>0.9201</td>
</tr>
<tr>
<td>0.92</td>
<td>0.6480</td>
<td>-0.3831</td>
<td>0.8419</td>
</tr>
<tr>
<td>0.93</td>
<td>0.5763</td>
<td>-0.3149</td>
<td>0.7856</td>
</tr>
<tr>
<td>0.94</td>
<td>0.5531</td>
<td>-0.2241</td>
<td>0.6808</td>
</tr>
<tr>
<td>0.95</td>
<td>0.4043</td>
<td>-0.1489</td>
<td>0.5845</td>
</tr>
<tr>
<td>0.96</td>
<td>0.5577</td>
<td>0.1682</td>
<td>0.4240</td>
</tr>
<tr>
<td>0.97</td>
<td>0.8762</td>
<td>0.7492</td>
<td>0.2245</td>
</tr>
<tr>
<td>0.98</td>
<td>0.9567</td>
<td>1.0018</td>
<td>0.1445</td>
</tr>
</tbody>
</table>

C. Performance Evaluation

In the following, we evaluate the efficiency of the above rate-curve decomposition scheme. For each sample picture, the estimated bit rate at $\rho_i$ is given by

$$\hat{R}_{H}(\rho_i) = \xi_{1}(\rho_i) \cdot Q_{cw}(\rho_i) + \xi_{2}(\rho_i) \cdot Q_{c}(\rho_i) + \xi_{3}(\rho_i).$$  \hspace{1cm} (11)

The relative estimation error in percentage is

$$\mathcal{E}(\rho_i) = \frac{R_{H}(\rho_i) - \hat{R}_{H}(\rho_i)}{R_{H}(\rho_i)} \times 100\%.$$  \hspace{1cm} (12)

Let

$$\mathcal{P}(\alpha, \rho_i) = \text{Probability}\{\mathcal{E}(\rho_i) < \alpha\}. \hspace{1cm} (13)$$

Table IV shows the values of $\mathcal{P}(3\%, \rho_i)$, $\mathcal{P}(4\%, \rho_i)$, $\mathcal{P}(5\%, \rho_i)$, and $\mathcal{P}(6\%, \rho_i)$ for all $\rho_i$. It can be seen that the probability of the relative estimation error being less than 5% is about 0.98. This implies that the approximation error given by (10) is very small. The above rate-decomposition scheme can also be applied to other video coding algorithms, such as MPEG-2 [2]. Certainly, different coding algorithms have different decomposition coefficients $\xi_{1}(\rho_i)$, $\xi_{2}(\rho_i)$, and $\xi_{3}(\rho_i)$. Note that, for a given picture, its rate function is only determined by the coding algorithm. From (10), we can see that the coding algorithm is modeled by $\{\xi_{1}(\rho_i), \xi_{2}(\rho_i), \xi_{3}(\rho_i)\}$, while the input picture is characterized by $Q_{cw}(\rho_i)$ and $Q_{c}(\rho_i)$. As mentioned in Section I,
the performance of a coding system is determined by two components: the characteristics of the input picture and the coding algorithm. We can see that (10) has naturally integrated these two components by linear combination.

IV. RATE-CONTROL ALGORITHM FOR H.263 VIDEO CODING

In the previous sections, we have developed a source modeling framework which can be applied to estimate the R-Q function for the H.263 video encoder. Based on the forward estimation of the R-Q function, the output bit rate and the buffer level of the video encoder can be well controlled. In this section, we first summarize the R-Q curve estimation algorithm. Based on this estimation algorithm, we then propose a frame-level rate control algorithm for H.263 video coding.

A. R-Q Curve Estimation Algorithm

Based on the discussions in Sections II and III, an R-Q curve estimation algorithm for DCT video coding is proposed as follows.

Step 1) Generate the distributions \(D_0(x)\) and \(D_2(x)\) for the DCT coefficients from the intra-coded and inter-coded macroblocks, respectively.

Step 2) Choose a quantization parameter \(q_0\) such that the percentage of zeros \(\rho\) is inside the interval \([0.9, 0.95]\). This restriction is imposed to make sure that the following computation is robust. With (5)–(7), compute the slope \(\theta\) and construct the whole rate curve \(Q_{m}(\rho)\) using (2). Compute the values of \(\{Q_{m}(\rho_{i})\}_{1 \leq i \leq 9}\) as indicated by (8). For \(1 \leq i \leq 9\), estimate \(R_{H}(\rho_{i})\) using (11).

Step 3) The one-to-one mapping between the quantization parameter \(q\) and the percentage of zeros \(\rho\) is computed with (6). Based on this mapping, \(\{R_{H}(\rho_{i})\}_{1 \leq i \leq 9}\) are mapped into the \(q\) domain. The whole R-Q curve is constructed by linear interpolation.

According to our experience, the above estimation algorithm can estimate the R-Q curve up to 1.5 bits per pixel (bpp) for H.263 video coding, which is quite sufficient in practical video coding. We can see that the proposed algorithm has very low computational complexity. The major complexity is the generation of the distributions of the DCT coefficients. The remaining computations only involve addition and multiplication operations, which are carried over the distributions. In addition, the R-Q curve is estimated before quantization and coding.

Compared to the rate curve, the distortion curve is much easier to estimate because it is determined only by the quantizer. Note that each coefficient is quantized independently. We can compute the distortion directly from the distributions of the DCT coefficients.

B. Rate Control Scheme for H.263 Video Coding

In a typical rate control scheme, the available bits for encoding the current video frame is obtained from the channel bandwidth and the buffer status [5]. The rate control algorithm can operate at two basic levels: macroblock and frame level. The TMN8 algorithm [5] employs a macroblock-level rate control scheme. The parameters of its source model are updated after each macroblock is encoded. The quantization parameter is determined at the macroblock level. In this paper, we already have a forward estimation of the R-Q curve for the current frame. Therefore, our rate control can be carried out at the frame level. This implies that the rate control needs to run only once per frame, which reduces the computational complexity and implementation cost.

Let \(C\) and \(F\) be the channel and frame rates, respectively. Let the buffer size be \(M\) which is set to be \(C/F\) by default in this work. If the number of bits in the buffer, denoted by \(W\), is larger than \(M\), the encoder then skips encoding frames until \(W\) is less than \(M\). For each skipped frame, the buffer fullness \(W\) is reduced by \(C/F\) bits. The target number of bits \(R_T\) for the encoding frame is determined as follows:

\[
R_T = \frac{C}{F} - \Delta
\]

where

\[
\Delta = W - \beta \cdot M.
\]

If we set \(\beta = 0.1\), with the above feedback control, the target buffer fullness is 10% of the buffer size. It is necessary to maintain a small number of bits in the buffer to avoid possible buffer underflow [5]. Let \(R_C\) be the available bits for encoding the DCT coefficients in the current frame which excludes the bits for coding the motion vectors and other header information. With the proposed R-Q estimation algorithm, the rate curve \(R_H(q)\) for the current frame can be estimated. Based on this curve, the frame quantization parameter, denoted by \(\tilde{Q}\), can be immediately determined to meet the target bits \(R_C\). The parameter \(\tilde{Q}\) is then applied to all the macroblocks in the current frame. It can be seen that the proposed rate control algorithm does not have a scene-change problem because it performs R-D estimation and control for each video frame independently from the others.

C. Quantization Parameter of Each MB

It should be noted that, in the H.263 encoder, the quantization parameter should have an integer value. However, the estimated quantization parameter \(\tilde{Q}\) is a real number. For example, suppose \(\tilde{Q} = 5.3\); the H.263 codec could round it to the integer 5. However, the respective coding bit rates of 5.3 and 5 are different. Certainly, this will affect our rate control accuracy. There are two basic approaches to solve this problem. In the first approach, we can apply the quantization parameter 6 to 30% of the macroblocks while applying 5 to the remaining macroblocks. In this way, the mean quantization parameter will be approximately 5.3. This approach is very simple. However, its rate control performance is not not very robust, since it does not consider the activities in the macroblocks.

The second approach employs a macroblock-level adaptive scheme. Suppose that, in the current frame, \(N_+\) macroblocks have already been encoded. Let \(\rho_+\) be the percentage of zeros produced by these macroblocks. Let \(R_+\) be the coding bit rate of the DCT coefficients in these macroblocks. The bit budget available for the remaining macroblocks is then \(R_C - R_+\). The per-
Fig. 6. (a)–(d) Comparison of the number of bits in the encoder buffer when the proposed rate control algorithm (solid line) and the TMNB (dashed line) are employed in the H.263 video coder. The name of respective video sequence, the channel rate $C$, and the frame rate $F$ are indicated in the title of each plot. The horizontal dashed-dotted line shows the buffer size $M$. The first frame is an I-frame which needs more coding bits than the P-frames.

The percentage of zeros to be produced by the remaining macroblocks is given by

$$\rho_\pm = R_H^{-1}(R_C - R_+),$$

where $R_H^{-1}(\cdot)$ represents the inverse function of the estimated rate function $R_H(\rho)$. We define

$$\kappa = \frac{\rho_-}{\rho_+},$$

and let (18), shown at the bottom of the page, where

$$Q_0 = \lceil Q \rceil.$$  \hspace{1cm} (19)

Here, $Q$ are six candidate quantization parameters for each macroblock in the current video frame. Suppose that we are now encoding the $(N_+ + 1)$-th macroblock. If $R_+ \leq 0.2 \cdot R_C$, the quantization parameter of the current macroblock $Q$ is set to be $Q_0$. Otherwise, we choose

$$Q = \begin{cases} Q_0 - 3, & \text{if } 1.3 \leq \kappa < \infty \\ Q_0 - 2, & \text{if } 1.2 \leq \kappa < 1.3 \\ Q_0 - 1, & \text{if } 1.1 \leq \kappa < 1.2 \\ Q_0, & \text{if } 0.9 \leq \kappa < 1.1 \\ Q_0 - 2, & \text{if } 0.8 \leq \kappa < 0.9 \\ Q_0 - 3, & \text{if } 0.7 \leq \kappa < 0.8 \\ Q_0 - 3, & \text{if } 0.0 \leq \kappa < 0.7. \end{cases}$$

$$Q = \begin{cases} Q_0 - 3, & \text{if } 1.3 \leq \kappa < \infty \\ Q_0 - 2, & \text{if } 1.2 \leq \kappa < 1.3 \\ Q_0 - 1, & \text{if } 1.1 \leq \kappa < 1.2 \\ Q_0, & \text{if } 0.9 \leq \kappa < 1.1 \\ Q_0 - 2, & \text{if } 0.8 \leq \kappa < 0.9 \\ Q_0 - 3, & \text{if } 0.7 \leq \kappa < 0.8 \\ Q_0 - 3, & \text{if } 0.0 \leq \kappa < 0.7. \end{cases}$$

(18)
It can be seen that our basic adaptive control policy is as follows. When we find that more zeros need to be produced in the remaining macroblocks, we use larger quantization parameter in \( Q \); otherwise, we use smaller one in \( Q \). Our experimental results in Section VI shows that this simple approach is very effective.

V. ENCODER-BASED RATE-SHAPE SMOOTHING

In VBR video coding, the output bit stream often has relative large rate fluctuations, which lead to packet loss and delay variation [9], [19]. From the standpoint of transmission efficiency, network traffic management, and resource allocation, VBR video is much harder to handle than the CBR video stream [19]–[21]. However, CBR coding often has a large variation in picture quality. Therefore, practical video encoders often operate between the CBR and VBR coding.

One method to smooth the frame bit rates without severe degradation of the picture quality is to introduce buffers on the packet delivery path [19]. The buffer operates like a low-pass filter. To achieve a good smoothing effect, the buffer size should be relatively large; e.g., 10–30 frames. This obviously will increase the implementation and traffic management cost. In addition, large buffer size results in large playback start-up latency and end-to-end delay. In this work, based on the estimated R-D functions, we propose an encoder-based rate-shape smoothing algorithm for real-time video coding and transmission. In this algorithm, a tradeoff between the CBR and VBR video coding is made at the encoder to smooth the rate shape at the cost of small variation of picture quality.

In real-time playback of coded video, a constant presentation quality throughout the whole video sequence is highly desirable. Let the target picture quality be \( D_T \) (in decibels). Let the bit rate of each video frame be \( R(n) \). Obviously, \( R(n) \) varies over time or is even bursty. Note that in practice, we do not have to keep the picture quality as a constant. Normally, the user can not tell the difference between the picture quality of \( D_T \) dB and \( D_T \pm 1 \) dB when the \( D_T \) is relative high, such as 35 or more. Based on this observation, we can let the picture quality of each frame, denoted...
by \( D(n) \), vary within a small range (such as \([D_T - 1, D_T + 1]\)) around the target picture quality \( D_T \). Obviously, this guarantees there is no severe video quality degradation over time.

With the small relaxation of \( D(n) \), we can now smooth the rate shape \( \{R(n)\} \) of the video stream. The smoothing algorithm is summarized as follows.

Step 1) **Initialization:** When \( n = 0 \), let \( R(0) \) be the bit rate such that \( D(0) = D_T \). The proposed rate control algorithm is then employed to achieve the target rate \( R(n) \).

Step 2) **Smoothing:** Suppose the current frame number is \( n \) (\( n \geq 1 \)). Let \( R_+ \) and \( R_- \) be the bit rates such that the corresponding picture quality \( D(n) \) is \( D_T + 1 \) and \( D_T - 1 \), respectively. Obviously, \( R_+ > R_- \). If \( R(n-1) \leq R_- \), we set \( R(n) = R_- \). If \( R(n-1) \geq R_+ \), we set \( R(n) = R_+ \). Otherwise, we set \( R(n) = R(n-1) \). This implies that we are trying to make \( R(n) \) as close to \( R(n-1) \) as possible. However, in any case, we always have \( D_T - 1 \leq D(n) \leq D_T + 1 \). The proposed rate control algorithm is then employed to achieve the target rate \( R(n) \).

This encoder-based smoothing algorithm imposes very little additional complexity to the proposed rate control algorithm. Our simulation results in Section VI shows that, with the proposed smoothing algorithm, we can obtain a video stream which has both a smoothed rate shape and a small variation (only \( \pm 1 \) dB) in picture quality. The rate-shape smoothing is a unique feature of the proposed source model and rate control algorithm. This is because the proposed source model can accurately predict the R-D functions of the video encoder.

**VI. EXPERIMENTAL RESULTS**

The proposed rate control algorithm is implemented in the UBC’s H.263 codec\(^1\) and tested for various video sequences and applications. The frame rate is fixed at 10 fps. All the test video sequences are in the QCIF picture format [1]. In each test described in the following, the name of the test video sequence, the channel rate \( C \), and the frame rate \( F \) are indicated in the title of the respective plot. In the following experiments on rate control, we compare the proposed algorithm with the TMN8 rate

\(^1\)The H.263+ codec is available at http://www.ece.ubc.ca/

---

**TABLE V**

<table>
<thead>
<tr>
<th>Test Video</th>
<th>Channel Rate</th>
<th>Frame Skipped</th>
<th>PSNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>This Work</td>
<td>TMN8</td>
<td>This Work</td>
</tr>
<tr>
<td>Foreman</td>
<td>64</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Foreman</td>
<td>48</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>M.c.D.</td>
<td>16</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Salesman</td>
<td>32</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Carphone</td>
<td>32</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Coastguard</td>
<td>32</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>News</td>
<td>48</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Container</td>
<td>32</td>
<td>1</td>
<td>4</td>
</tr>
</tbody>
</table>
control algorithm [5] which performs much better than other algorithms reported in the literature, such as VM7 [6].

A. Buffer Regulation Performance

In Figs. 6(a)–(h), we plot the number of bits in the encoder buffer for each coded frame when the proposed rate control algorithm and the TMN8 algorithm are applied. The horizontal dashed-dotted line shows the buffer size $M$. This is the threshold for frame skipping. Note that the first frame in each video is an I-frame, which produces a lot of bits. As a result, the buffer overflows and several frames are skipped. When there are fewer bits left in the buffer than the channel rate, the buffer underflows. For example, this happens at frame 153 in Fig. 6(g) and frame 270 in Fig. 6(h) when the TMN8 algorithm is applied. When buffer underflows, a part of the channel bandwidth is wasted. From these experimental results, we can see that the proposed frame-level rate control algorithm is competitive to, or even outperforms the TMN8 algorithm, especially at low bit rates.

To test the performance of the proposed algorithm at scene changes, we first generate a combination video by concatenating the first 30 frames of following five videos: Foreman, Carphone, Salesman, Miss America, and Coastguard. In this way, a scene change occurs between any two neighboring video clips. In Fig. 7, we plot the buffer fullness when the proposed rate control algorithm and the TMN8 are applied. Again, the proposed algorithm is seen to work very well at scene changes.

In Fig. 8, we plot the number of bits produced by each coded frame when the proposed and the TMN8 algorithms are applied. The proposed algorithm maintains a output bit rate closer to the channel rate $C$. In Table V, we compare the average PSNR and the number of the skipped frames for the proposed and the TMN8 algorithms. We can see that the new algorithm has fewer skipped frames and better PSNR performance. The improved picture quality of our algorithm is due to its accurate R-D models and robust rate control.

B. Dynamic Range of the Quantization Parameters

In TMN8, QP is determined at the macroblock level. There is no restriction for the range of QP. However, in the proposed rate control algorithm, based on the estimated R-Q curve of the

Fig. 9. Comparison of the range of the quantization parameters in each frame. The test video sequence are (a) “Coastguard” with a channel rate of 32 kbit/s; (b) “Carphone” with a channel rate of 24 kbit/s; (c) “Mother and Daughter” with a channel rate of 24 kbit/s; and (d) “Salesman” with a channel rate of 32 kbit/s.
Fig. 10. Reconstructed 114th frame in “Carphone” when: (a) the proposed algorithm is used and (b) the TMN8 is used. The channel rate is 24 kbits/s.

Fig. 11. Reconstructed 270th frame in “Mother and Daughter” when: (a) the proposed algorithm is used and (b) the TMN8 is used. The channel rate is 10 kbits/s.

current frame, we first determine the mean quantization parameter $\bar{Q}$ for the whole frame. The quantization parameter $QP$ for each macroblock is then chosen from the neighborhood of $\bar{Q}$. Therefore, the quantization parameter for each macroblock is well controlled.

Let $QP_{\text{max}}$ be the maximum and $QP_{\text{min}}$ be the minimum macroblock quantization parameters in the current frame. The dynamic range of $QP$ is then given by $QP_{\text{max}} - QP_{\text{min}}$. (Obviously, the range of 0 implies that all of the macroblocks in the current frame use the same $QP$.) In Figs. 9(a)–(d), we plot the dynamic ranges of $QP$ for each frame when the proposed algorithm and the TMN8 are applied to H.263 video coding. The maximum range of the proposed algorithm is 6. However, in the TMN8 rate control algorithm, the range could be very large, perhaps as large as 29. Note that in H.263 coding, the maximum quantization parameter is 31. This implies that in TMN8, some blocks in the current frame are quantized by extremely small QPs, while others are quantized by extremely large QPs. As a result, the quality inside each picture has a relative large variation. Hence, the overall subjective quality degrades. To show this, let us examine the following two video frames: the 114th frame in “Carphone” and the 270th frame in “Mother and Daughter.” The reconstructed frames by the proposed rate control algorithm are shown in Figs. 10(a) and 11(a), respectively. The reconstructed frames by the TMN8 algorithm are shown in Figs. 10(b) and 11(b). It can be seen that the proposed algorithm produces a better picture quality even though the PSNR values are almost the same.

The above phenomenon is due to the inherent difference between the two rate control algorithms. In the TMN8 algorithm, since the source model is not very accurate, the model parameters are empirically adjusted during the encoding process. In addition, this type of empirical adjustment is not regulated [5]. Another disadvantage of a very large dynamic range of $QP$ is that some very important information in the picture might be lost. For example, suppose the region of interest is at the bottom half of the picture. The TMN8 rate control algorithm uses very large $QP$ to quantize this part. Its quality will be poor and the valuable information will be lost. However, in the proposed rate control algorithm, since the quantization parameter for each macroblock is restricted inside a small range, a steady and uniform picture quality is guaranteed.

C. Adaptivity to the Channel Bandwidth

In live video coding and transmission, the channel bandwidth varies over time. Normally, the video encoder can detect the available channel bandwidth from the feedback information. To simulate the rate control for live video coding in this situation, we assume that the channel bandwidth is varying as plotted in Fig. 12 (top). The frame bits for “News” at 24 kbits/s are plotted in Fig. 12 (bottom) when the proposed algorithm and the TMN8 rate control algorithm are applied. It can be seen that with the proposed rate control algorithm, the encoder can better match the coding bit rate to the channel bandwidth.
Fig. 12. Comparison of the rate control performance with varying channel bandwidth for “News” at 24 kbits/s.

Fig. 13. Rate shapes with and without smoothing for “Foreman” when the picture quality variation range is 1 dB.

D. Rate-Shape Smoothing

As discussed in Section V, the proposed rate control algorithm can maintain a video stream with both a smoothed rate shape and a small variation of picture quality. In the following experiment, the test video is “Foreman.” If we set the target picture quality for each frame to 34 dB. The corresponding rate for each frame is plotted in Fig. 13 as a dotted line. It can be seen that the rate varies dramatically over time. If we apply the proposed rate-shape smoothing algorithm, the smoothed rate is shown in Fig. 13 as a solid line. We can see that after smoothing, the rate burtness is significantly reduced. The corresponding picture quality of each frame is plotted in Fig. 14. It can be seen that always falls in a small range between 33 and 35 dB, which implies that we have a consistent picture quality over the playback time. In the above, the picture quality relaxation is 1.0 dB. If we change it to 2 dB, the corresponding rate shape is even smoother, as shown in Fig. 15. This unique feature of the proposed rate control algorithm is highly desirable in real-time video coding and transmission.

VII. CONCLUDING REMARKS

We have presented the \( \rho \)-domain analysis, a novel and efficient method for studying the R-D behavior of a transform-coding system. Based on this methodology, by introducing the new concepts of characteristic rate curves and rate-curve decomposition, a unified source-modeling framework is developed. Within this framework, an accurate and forward R-D estimation and control algorithm is proposed for DCT video coding. Our extensive experimental results show that the proposed algorithm outperforms the rate control algorithm reported in the literature by providing more accurate rate regulation and better picture quality. Based on the estimated R-D curves, an encoder-based rate-shape-smoothing algorithm is proposed. Using this smoothing algorithm, we can control the encoder such that the output video stream has both a smoothed rate shape and a consistent picture quality, which are highly desirable in real-time video coding and transmission.

REFERENCES


Zhizhai He (M’01) received the B.S. degree from Beijing Normal University, Beijing, China, in 1994, and the M.S. degree from the Institute of Computational Mathematics, Chinese Academy of Sciences, Beijing, China, in 1997, both in mathematics, and the Ph.D. degree in electrical engineering from the University of California at Santa Barbara in 2001. He joined Sarnoff Corporation, Princeton, NJ, as a Member of Technical Staff. His current research interests include video coding and communication.

Yong Kwan Kim (S’88–M’97) was born in Seoul, Korea, in 1965. He received the B.S., M.S., and Ph.D. degrees in control and instrumentation engineering from Seoul National University, Seoul, Korea, in 1988, 1990, and 1996, respectively. During 1994–1997, he was with Daewoo Electronics Company, Ltd., working on the development of HDTV systems. In 1997, he joined the Department of Information and Telecommunication Engineering, Hoseo University, Chungnam-do, Korea, where he is currently an Assistant Professor. He was a visiting scholar in electrical and computer engineering at the University of California at Santa Barbara during 2000–2001. His current research interests include image processing, video compression, video transmission systems, and multimedia applications.

Sanjit K. Mitra (S’59–M’63–SM’69–F’74–LF’00) received the B.Sc. (Hons.) degree in physics in 1953 from Utkal University, Cuttack, India, the M.Sc. (Tech.) degree in radio physics and electronics from Calcutta University, Calcutta, India, in 1956, the M.S. and Ph.D. degrees in electrical engineering from the University of California at Berkeley in 1960 and 1962, respectively, and an Honorary Doctorate of Technology degree from the Tampere University of Technology, Tampere, Finland.

From 1962 to 1965, he was with Cornell University, Ithaca, NY, as an Assistant Professor of Electrical Engineering. He was with the AT&T Bell Laboratories, Holmdel, NJ, from June 1965 to January 1967. He has been on the faculty of the University of California since 1967, serving as a Professor of Electrical and Computer Engineering since 1977 and Chairman of the Department from July 1979 to June 1982. He has published over 500 papers on signal and image processing, 11 books, and holds five patents. Dr. Mitra served as the President of the IEEE Circuits and Systems (CAS) Society in 1986 and as a Member-at-Large of the Board of Governors of the IEEE Signal Processing (SP) Society from 1986-1999. He is currently a member of the editorial boards of *Multidimensional Systems and Signal Processing, Signal Processing, Journal of the Franklin Institute, and Automatica*. He is the recipient of numerous awards, including the 1973 F.E. Terman Award, the 1985 AT&T Foundation Award of the American Society of Engineering Education, the Education Award of the IEEE CAS Society in 1989, the Distinguished Senior U.S. Scientist Award from the Alexander von Humboldt Foundation of Germany in 1989, the Technical Achievement Award of the IEEE SP Society in 1996, the Mac Van Valkenburg Society Award, and the CAS Golden Jubilee Medal of the IEEE CAS Society in 1999, and the IEEE Millennium Medal in 2000. He is an Academician of the Academy of Finland, a Fellow of the AAAS and SPIE, and a member of EURASIP and ASEE.