A Unified Rate-Distortion Analysis Framework for Transform Coding

Zhihai He, *Member, IEEE,* and Sanjit K. Mitra, *Life Fellow, IEEE*

***Abstract*—In our previous work, we have developed a rate-distortion (R-D) modeling framework H.263 video coding by introducing the new concepts of characteristic rate curves and rate curve decomposition. In this paper, we further show it is a unified R-D analysis framework for all typical image/video transform coding systems, such as EZW, SPIHT and JPEG image coding; MPEG-2, H.263, and MPEG-4 video coding. Based on this framework, a unified R-D estimation and control algorithm is proposed for all typical transform coding systems. We have also provided a theoretical justification for the unique properties of the characteristic rate curves. A linear rate regulation scheme is designed to further improve the estimation accuracy and robustness, as well as to reduce the computational complexity of the R-D estimation algorithm. Our extensive experimental results show that with the proposed algorithm, we can accurately estimate the R-D functions and robustly control the output bit rate or picture quality of the image/video encoder.**

***Index Terms*—Rate control, rate-distortion analysis, source modeling, transform coding, video coding and transmission.**

I. Introduction

RECENT advances in computing and communication technology have stimulated the research interest in digital techniques for recording and transmitting visual information. The exponential growth in the amount of visual data to be stored, transferred, and processed has created a huge need for data compression. Compression of visual data, such as images and videos, can significantly improve the utilization efficiency of the limited communication channel bandwidth or storage capacity.

A. Transform Coding

The demand for image and video compression has triggered the development of several compression standards, such as JPEG [2], JPEG-2000 [3], MPEG-2 [4], H.263 [5], and MPEG-4 [6]. Besides the standard image/video compression algorithm, many other algorithms have also been reported in the literature, such as embedded zero-tree wavelet (EZW) [7] image coding, set partitioning in hierarchical trees (SPIHT) [8] and stack-run (SR) [9] image coding. In both the compression standards and the algorithms reported in the literature, transform coding has become the dominant approach for image and video compression. A generic transform coding system is depicted in Fig. 1. The transform, either discrete wavelet transform (DWT) or discrete cosine transform (DCT), is applied to the input picture. Here, a picture can be either a still image or motion-compensated video frame. After quantization, the quantized coefficients are converted into symbols according to some data representation scheme. For example, zig–zag scan and run-level data representation are employed in JPEG and MPEG coding [2], [4]. In embedded zero-tree wavelet (EZW) coding [7], all insignificant coefficients in a spatial orientation tree are represented by one zero-tree symbol. After data representation, the output symbols are finally encoded by a Huffman or arithmetic coder [13].

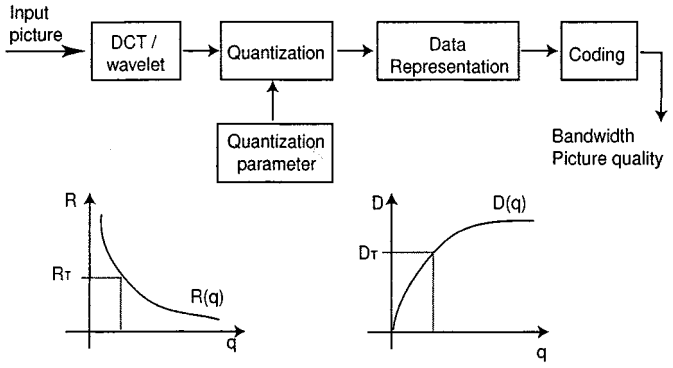

Fig. 1. Generic transform coding system for images and videos.

B. R-D Analysis

In transform coding of images and videos, the two most important factors are the coding bit rate and picture quality. The coding bit rate R determines the channel bandwidth required to transfer the coded visual data. One direct and widely used measure for the picture quality is the mean-square error (MSE) between the coded image/video and the original one. The reconstruction error introduced by compression, often referred to as distortion, is denoted by D. In typical transform coding, both R and D are controlled by the quantization parameter of the quantizer q. The major issue here is how to determine the value of q to achieve the target coding bit rate R_T, or target picture quality D_T. To this end, we need to analyze and estimate the R-D behavior of the image/video encoder; this behavior is characterized by its rate-quantization (R-Q) and distortion-quantization (D-Q) functions, $R(q)$ and $D(q)$, respectively [10], [11]. In this work, they are collectively called *R-D functions* or *curves*. Based on the R-D functions, the quantization parameter q can be readily determined to achieve the target bit rate R_T or picture

Manuscript received May 8, 2000; revised September 28, 2001. This paper was recommended by Associate Editor S. U. Lee.

Z. He was with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106 USA. He is now with the Interactive Media Group, Sarnoff Corporation, Princeton, NJ 08543-5300 USA (e-mail: zhe@sarnoff.com).

S. K. Mitra is with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106 USA (e-mail: mitra@ece.ucsb.edu).

Publisher Item Identifier S 1051-8215(01)11031-1.

quality D_T [12], [14]. Therefore, the major issue becomes this: how to analyze, model and estimate the R-D functions for the image/video encoder.

Analysis and estimation of the R-D functions have important applications in visual coding and communication. First, with the estimated R-D functions we can adjust the quantization setting of the encoder and control the output bit rate or picture quality according to the channel condition, the storage capacity, or the user's requirement [14]–[16]. Second, based on the estimated R-D functions, optimum bit allocation as well as other R-D optimization procedures can be performed to improve the efficiency of the coding algorithm and, consequently, to improve the image quality or video presentation quality [17], [18].

There are two basic approaches for R-D modeling. The first is the analytic approach. Its objective is to derive a set of mathematical formulas for the R-D functions based on the statistical properties of the source data. In this approach, both the coding system and the image are first decomposed into components whose statistical models are already known. These models are then combined to form a complete analytic model for the whole coding system. The R-D functions for a simple quantizer have been developed for a long time [10], [11]. In the analytic source model proposed by Hang and Chen [12], a theoretical entropy formula for the quantized DCT coefficients is developed based on the R-D theory of the Gaussian source and the uniform quantizer. The mismatches between the theoretical entropy and the actual coding bit rate of the entropy encoder is, however, compensated by empirical estimation.

The second approach is the empirical approach. Here, the R-D functions are constructed by mathematical processing of the observed R-D data. In the R-D estimation algorithm proposed by Lin and Ortega [23], eight control points on the R-D curves are first computed by running the coding system eight times. The whole R-D curves are then constructed by cubic interpolation. In the MPEG rate control algorithm proposed by Ding and Liu [15], the R-D curves are fitted by mathematical functions with several control parameters which are estimated from the observed R-D data of the coding system. In general, this type of R-D estimation algorithms have very high computational complexity. In addition, such algorithms do not provide us with insights into the R-D behaviors of the transform coding systems

Within the context of video coding, the coding results of the previous frames or macroblocks (MBs) can be utilized to estimate the R-D functions of the current frame or MB. This adaptive estimation scheme is employed in many rate control algorithms proposed in the literature, such as the MPEG-2 Test Model Version 5 (TM5) rate control algorithm [19], the H.263 Test Model Near-term Version 8 (TMN8) algorithm [20], and the MPEG-4 Verification Model Version 8 (VM8) algorithm [16], [21]. These rate control algorithms often suffer from relative large control errors and perform degradation at scene changes, especially at low bit rates for active videos.

C. Proposed R-D Analysis Framework

It is well known that the R-D behavior of an image/video encoder is determined both by the characteristics of the input source data and by the capability of the coding algorithm to explore these characteristics. The R-D models reported in the literature try to use some statistics of the input source data, such as variance, to describe the input image or video data [10], [12], [14]. They also try to analyze and model each step of the coding algorithms and formulate an explicit expression of the coding bit rate. To achieve high coding performance, an efficient coding algorithm must employ a sophisticated data representation scheme as well as an efficient entropy coding scheme. To improve the rate estimation accuracy for these coding algorithms, the rate models are becoming very complex [12], [14], [24], [25]. However, with complex and highly nonlinear expressions, the estimation and rate control process becomes increasingly complicated and even unstable with the image-dependent variations [23].

It should also be noted that, for different coding algorithms, the R-D models and rate control algorithms reported in the literature are quite different from each other [12], [14]–[16], [24], [25]. It would be ideal to develop a simple, accurate, and unified rate model for all typical transform coding systems. Based on this simple model, we could then develop a unified rate and picture quality control algorithm which could be applied to all typical transform coding systems. To this end, we need to uncover the common rules that govern the R-D behaviors of all transform coding systems. Obviously, this will provide us with valuable insights into the mechanism of transform coding. From a practical point of view, the simple and unified rate model and control algorithm would enable us to control the image/video encoder accurately and robustly with very low computational complexity and implementation cost.

In this work, based on the so-called ρ-domain analysis method proposed in [1], [22], we develop a generic source modeling framework for transform coding of images and videos by the following two major steps. In the first step, we introduce the concepts of characteristic rate curves and rate curve decomposition to characterize the input source date and to model the coding algorithm, respectively. In the second step, we propose a linear regulation scheme to improve the accuracy and robustness of the R-D estimation. Our extensive simulation results show that the proposed framework is a unified R-D analysis framework for all typical transform coding, such as EZW, SPIHT, and JPEG image coding and MPEG-2, H.263, and MPEG-4 video coding. With the estimated R-D functions, the output bit rate and picture quality of the image/video encoder can be controlled in flexible way according to the requirements imposed by the specific applications.

D. Paper Organization

The remainder of this paper is organized as follows. In Section II, we generalize the ρ-domain R-D analysis methodology for all typical transform coding system. Section III outlines the proposed source modeling framework. In Section IV, we define the characteristic rate curves and show their unique properties. Based on these properties, we then propose a fast algorithm to estimate them with very low computational complexity. The rate curve decomposition scheme is explained in Section V. A rate regulation scheme is proposed in Section VI to improve the R-D estimation accuracy and robustness. The R-D estimation algorithm is summarized in Section VII. In Section VIII, we present

the R-D estimation and control results for still image and video coding. Concluding remarks are provided in Section IX.

II. ρ-Domain R-D Analysis

It has been observed that zeros play a key role in transform coding, especially at low bit rates. All typical coding algorithms treat zeros in a special way and address most of the effort to efficient coding of them. For example, in JPEG and MPEG coding, run-length representation and a special symbol of end-of-block (EOB) are employed to code the zeros [2], [4]. In H.263 video coding, a special binary flag named "LAST" is introduced to signal that all the remaining coefficients in a zig–zag order in the current block are zeros [5]. After the DCT coefficients are quantized with a quantization parameter q, let ρ be the percentage of zeros among the quantized coefficients. Note that in typical transform coding systems, ρ monotonically increases with q. (Here, we have made a trivial assumption that the distribution of the transform coefficients is continuous and positive.) Hence, there is a one-to-one mapping between them. This implies that, mathematically, R and D are also functions of ρ, denoted by $R(\rho)$ and $D(\rho)$. A study of the rate and distortion as functions of ρ is called *ρ-domain analysis*.

A. Typical Quantization Schemes

To map the R-D functions between the q- and ρ-domains, we first need to obtain the one-to-one mapping between q and ρ. Note that this mapping depends on the quantization scheme. In the following, we briefly review the quantization schemes employed by the typical transform coding systems before discussing the computation of the mapping between q and ρ.

1) Quantization in Wavelet Image Coding: In the typical wavelet-based image coding systems, uniform threshold quantization (UTQ) is often used, either explicitly or implicitly. In this case, the quantization parameter q refers to the UTQ stepsize. Let Δ be the UTQ dead zone threshold. In general, Δ is proportional to q. For any transform coefficient x, its UTQ output index is given by

$$I[x] = \mathrm{UTQ}[q, \Delta; x] = \begin{cases} 0, & \text{if } |x| \le \Delta \\ \lceil \frac{x-\Delta}{q} \rceil, & \text{if } x > +\Delta \\ \lfloor \frac{x+\Delta}{q} \rfloor, & \text{if } x < -\Delta. \end{cases} \quad (1)$$

2) H.263 Quantization Scheme: The quantization scheme employed by H.263 video coding is similar to UTQ. To be more specific, the quantization index of x in the H.263 style quantization scheme is given by (2), shown at the bottom of the page. Note that the range of the unquantized DC coefficient is 0–2048, which implies the range of its differential value is -2048 to 2048. In H.263 coding, it is quantized by a uniform quantizer with fixed step size 8, as shown in (2).

3) JPEG Quantization Scheme: In JPEG still image coding, a perceptual quantization scheme is employed. Each of the 64 DCT coefficients in an 8×8 block is quantized by a different uniform quantizer (UQ). The actual step sizes for the coefficients in the luminance component are associated with a quantization matrix, denoted by $[w_J^l(i,j)]_{1 \le i,j \le 8}$, where

$$[w_J^l(i,j)] = \begin{bmatrix} 16 & 11 & 10 & 16 & 24 & 40 & 51 & 61 \\ 12 & 12 & 14 & 19 & 26 & 58 & 60 & 55 \\ 14 & 13 & 16 & 24 & 40 & 57 & 69 & 56 \\ 14 & 17 & 22 & 29 & 51 & 87 & 80 & 62 \\ 18 & 22 & 37 & 56 & 68 & 109 & 103 & 77 \\ 24 & 35 & 55 & 64 & 81 & 104 & 113 & 92 \\ 49 & 64 & 78 & 87 & 103 & 121 & 120 & 101 \\ 72 & 92 & 95 & 98 & 112 & 100 & 103 & 99 \end{bmatrix}. \quad (3)$$

Let $x(i,j)$ be the DCT coefficient located at (i,j) inside a luminance block. Its quantization output is given by

$$I[x(i,j)] = \mathrm{Round}\left[\frac{x(i,j)}{q \cdot w_J^l(i,j)}\right] \quad (4)$$

where the quantization parameter q functions as a scaling factor which controls the coding bit-rate and the picture quality. If $x(i,j)$ is from a chrominance block, $w_J^l(i,j)$ in (4) is then replaced by the chrominance quantization matrix $[w_J^c(i,j)]$, where

$$[w_J^c(i,j)] = \begin{bmatrix} 17 & 18 & 24 & 47 & 99 & 99 & 99 & 99 \\ 18 & 21 & 26 & 66 & 99 & 99 & 99 & 99 \\ 24 & 26 & 56 & 99 & 99 & 99 & 99 & 99 \\ 47 & 66 & 99 & 99 & 99 & 99 & 99 & 99 \\ 99 & 99 & 99 & 99 & 99 & 99 & 99 & 99 \\ 99 & 99 & 99 & 99 & 99 & 99 & 99 & 99 \\ 99 & 99 & 99 & 99 & 99 & 99 & 99 & 99 \\ 99 & 99 & 99 & 99 & 99 & 99 & 99 & 99 \end{bmatrix}. \quad (5)$$

4) MPEG Quantization Scheme: In MPEG-2 coding, the JPEG-style perceptual quantization scheme is employed. The quantization matrices for intracoded and intercoded MBs, denoted by $w_M^0(i,j)$ and $w_M^1(i,j)$, respectively, are given as follows:

$$[w_M^0(i,j)] = \begin{bmatrix} 8 & 16 & 19 & 22 & 26 & 27 & 29 & 34 \\ 16 & 16 & 22 & 24 & 27 & 29 & 34 & 37 \\ 19 & 22 & 26 & 27 & 29 & 34 & 34 & 38 \\ 22 & 22 & 26 & 27 & 29 & 34 & 37 & 40 \\ 22 & 26 & 27 & 29 & 32 & 35 & 40 & 48 \\ 26 & 27 & 29 & 32 & 35 & 40 & 48 & 58 \\ 26 & 27 & 29 & 34 & 38 & 46 & 56 & 69 \\ 27 & 29 & 35 & 38 & 46 & 56 & 69 & 83 \end{bmatrix} \quad (6)$$

$$I[x] = \begin{cases} \mathrm{Round}\left(\frac{x}{8}\right), & \text{if } x \text{ is a DC coefficient in an intra-MB} \\ \mathrm{UTQ}(2q, 2q; x), & \text{if } x \text{ is an AC coefficient in an intra-MB} \\ \mathrm{UTQ}(2q, 2.5q; x), & \text{if } x \text{ is a coefficient in an inter-MB.} \end{cases} \quad (2)$$

$$[w_M^1(i,j)] = \begin{bmatrix} 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \end{bmatrix}. \tag{7}$$

Unlike the JPEG quantization, in the same MB, both the luminance and chrominance component use the same quantization matrix. In MPEG-2 coding, the quantization index of the DCT coefficient $x(i,j)$ is given by (8), shown at the bottom of the page. In MPEG-4 standard, both the H.263 style and the MPEG-2 style quantization are adopted. The user needs to configure the encoder to choose the quantization scheme. It can be seen that all the quantization schemes listed in the above are very close to the uniform threshold quantization.

B. The Mapping Between q and ρ

The one-to-one mapping between q and ρ can be directly computed from the distribution information of the transform coefficients. This is because in all typical transform coding systems each transform coefficient is quantized separately. In the following, we describe in detail how to compute the one-to-one mapping between q and ρ for different coding systems.

1) Wavelet Image Coding: The wavelet-based image coding schemes such as EZW and SR employ the uniform threshold quantization scheme given by (1), either explicitly or implicitly. Let the distribution of the wavelet coefficients be $\mathcal{D}(x)$. After quantization, the percentage of zeros among the quantized transform coefficients is given by

$$\rho(q) = \frac{1}{M}\int_{-\Delta}^{+\Delta} \mathcal{D}(x)dx \tag{9}$$

where M is the image size.

2) H.263 Coding: The H.263 quantization scheme is given by (2). Let $\mathcal{D}_0(x)$ and $\mathcal{D}_1(x)$ be the distributions of the DCT coefficients in the intracoded and intercoded MBs, respectively. Note that, in general, the DC coefficients from the intracoded MBs will not be quantized to zeros because of their relatively large values. Therefore, for any quantization parameter q, the corresponding percentage of zeros ρ can be obtained as follows:

$$\rho(q) = \frac{1}{M}\int_{-2q}^{+2q} \mathcal{D}_0(x)dx + \frac{1}{M}\int_{-2.5q}^{+2.5q} \mathcal{D}_1(x)dx \tag{10}$$

where M is the number of coefficients in the current video frame. Note that in the H.263 codec, the DCT coefficients are rounded to integers [27]. Therefore, $\mathcal{D}_0(x)$ and $\mathcal{D}_1(x)$ are actually histograms of the DCT coefficients, and (10) actually becomes

$$\rho = \frac{1}{M}\sum_{|x|<2q} \mathcal{D}_0(x) + \frac{1}{M}\sum_{|x|<2.5q} \mathcal{D}_1(x). \tag{11}$$

3) JPEG and MPEG Coding: Perceptual quantization is employed in the JPEG image coding, and in MPEG-2 and MPEG-4 video coding. Detailed descriptions are given in (4) and (8). After DCT, we first divide each DCT coefficient by its associated perceptual weight, then generate the distribution of these scaled DCT coefficients. After scaling, the perceptual quantization becomes uniform, as we can see from (4) and (8). Therefore, (9) and (11) can be also used to compute the value of ρ from q for JPEG and MPEG coding algorithms, respectively.

C. Implementation

From the distribution of the transform coefficients, for any given quantization parameter q we can compute the corresponding ρ. In software implementation, we can store the one-to-one mapping between q and ρ in a look-up table. For example, in H.263 and MPEG video coding, the possible values of q are $1, 2, \ldots, 31$. So, the look-up table has at most 31 entries. Using this look-up table, we can easily map the R-D functions between the q-domain and the ρ-domain.

III. Proposed R-D Modeling Framework

In Fourier analysis, which is a powerful tool for digital signal processing, to study the behavior of a function $f(x)$, we first represent $f(x)$ by a linear combination of the basis functions $\{\sin(nx), \cos(nx)\}$ which have well-known properties as follows:

$$f(x) = a_0 + \sum_{n=1}^{\infty} a_n \cos(nx) + b_n \sin(nx) \tag{12}$$

where $\{a_n, b_n\}$ are called Fourier coefficients. By studying these Fourier coefficients, we can then analyze the behavior of $f(x)$. This method is referred to as *signal decomposition and spectrum analysis*. In this chapter, we apply this decomposition scheme to analyze and estimate the rate function of an image/video encoder.

To estimate $R(\rho)$ using the decomposition scheme, we first define two basis functions $Q_{nz}(\rho)$ and $Q_z(\rho)$, called characteristic rate curves, to characterize the input source data. Here, the source data can be a still image or a video frame. We then show that, in the ρ-domain, $Q_{nz}(\rho)$ and $Q_n(\rho)$ have unique behaviors that enable us to estimate them with very low computational complexity. In our decomposition scheme, the actual rate

$$I[x(i,j)] = \begin{cases} \text{Round}\left(\frac{x}{8}\right), & \text{if } x \text{ is a DC coefficient in an intra-MB} \\ \text{Round}\left[\frac{16\cdot x(i,j)}{2\cdot q\cdot w_M^0(i,j)}\right], & \text{if } x \text{ is an AC coefficient in an intra-MB} \\ \text{Round}\left[\frac{16\cdot x(i,j)}{2\cdot q\cdot w_M^1(i,j)}\right], & \text{if } x \text{ is a coefficient in a nonintra MB.} \end{cases} \tag{8}$$

function is represented by a linear combination of $Q_{nz}(\rho)$ and $Q_n(\rho)$

$$R(\rho) = A(\rho) \cdot Q_{nz}(\rho) + B(\rho) \cdot Q_z(\rho) + C(\rho), \qquad (13)$$

where $A(\rho)$, $B(\rho)$, and $C(\rho)$ are the rate decomposition coefficients. For a given input image, $Q_{nz}(\rho)$ and $Q_z(\rho)$ are determined by their definitions. If we use different coding algorithms to encode this image, we should obtain different $R(\rho)$. According to (13), we know the corresponding decomposition coefficients should also be different. In other words, different coding algorithms correspond to different decomposition coefficients. Therefore, we can say that $\{A(\rho), B(\rho), C(\rho)\}$ model the coding algorithm, while $\{Q_{nz}(\rho), Q_z(\rho)\}$ characterize the input source data. As mentioned above, the R-D performance of a coding system is determined by these two components. We see that both of these components have been integrated by linear combination into (13), which serves as the framework for our ρ-domain source modeling. In Section IV, we define $Q_{nz}(\rho)$ and $Q_z(\rho)$, analyze their properties, and discuss the rate decomposition scheme in detail.

IV. Characteristic Rate Curves

In [1], we have defined two characteristic rate curves, $Q_{nz}(\rho)$ and $Q_z(\rho)$, to describe the input source data for H.263 video coding. In this work, we generalize their definitions for all typical image/video transform coding systems. Based on our extensive simulations, we then show that they have unique statistical properties, which hold for any of these coding systems. We also provide a theoretical justification for the unique properties. With these properties, a fast algorithm is proposed to estimate these two rate curves.

A. Definition

The characteristic rate curves $Q_{nz}(\rho)$ and $Q_z(\rho)$ are employed to characterize the transform coefficients to be quantized and coded by the image/video encoder. Their definitions are based on the binary representation of the nonzero coefficients and the run length numbers of zeros. This is because, as explained in [1], we believe that the binary representation collects the most valuable information about the R-D behavior of the transform coefficients. Based on the binary representation, we define $Q_{nz}(\rho)$ and $Q_z(\rho)$ as follows.

Step 1) *Conversion to 1-D array*: After transform and quantization with a quantization parameter q, the transform coefficients are rearranged into a 1-D array $\mathcal{L}$. If DWT is used, the subband coefficients are rearranged into $\mathcal{L}$ in a raster scan order. If DCT is used, all the DCT coefficients are rearranged into $\mathcal{L}$ in a zig–zag scan order inside each block and a block-wise raster scan order at the block level.

Step 2) *Binary Representation*: For any nonzero number x, its size $S(x)$ is defined as

$$S(x) = \lfloor \log_2 |x| \rfloor + 2 \qquad (14)$$

which is exactly the number of bits for its sign-magnitude representation. Note that, according to the above definition, $S(\pm 1)$ is 2 instead of 1. For each continuous string of zeros in $\mathcal{L}$, we count their run length. Let Q'_z be the sum of the sizes of all the run length numbers. For all the nonzero transform coefficients in $\mathcal{L}$, we define

$$Q'_{nz} = \sum_{x \in \mathcal{L}, x \neq 0} S(x) \qquad (15)$$

which is the sum of their sizes. Let

$$Q_{nz} = \frac{1}{M} Q'_{nz}, \quad Q_z = \frac{1}{M} Q'_z \qquad (16)$$

where M is the number of coefficients inside the picture. Q_{nz} and Q_z can be respectively regarded as the pseudo-coding bit rates for the nonzeros and zero coefficients. Obviously, they are functions of q. Let ρ be the percentage of zeros among the quantized transform coefficients. From Section II, we know that there is a one-to-one mapping between q and ρ. Therefore, mathematically, Q_{nz} and Q_z are also functions of ρ, denoted by $Q_{nz}(\rho)$ and $Q_z(\rho)$, respectively. These two functions are called the characteristic rate curves.

We would like to point out one implementation detail about this definition. In H.263 video coding, the DC coefficients from the intracoded MBs are quantized with a fixed quantization parameter 8 and encoded with a fixed number of bits which is also 8. This implies that the coding bit rate of these DC coefficients is fixed and does not depend on the quantization parameter. Therefore, when we scan the picture to form the 1-D array, the DC coefficients from the intracoded MBs are all skipped. However, their coding bit rate will be compensated by our rate decomposition scheme presented in Section V.

B. Statistical Properties

In [1], we observed that the two characteristic rate curves $Q_{nz}(\rho)$ and $Q_z(\rho)$ have unique properties for H.263 video coding. In this section, we show that these properties hold for all typical image/video transform coding systems.

1) Still Image Coding: For each sample image in Fig. 2, we first decompose it with a 9–7 Debauchies wavelet [26]. According to their definitions, we generate the rate curves $Q_{nz}(\rho)$ and $Q_z(\rho)$ and plot them in Fig. 3. Two observations can be made from these plots. First, although the sample images are quite different from each other, their characteristic rate curves share the same pattern. The second observation is that Q_{nz} is approximately a straight line. Note that when ρ is 1.0, which means all the coefficients are quantized to zeros, by definition Q_{nz} is 0, i.e., $Q_{nz}(\rho)$ must pass through the point $[1.0, 0.0]$. Hence, it has the following expression:

$$Q_{nz}(\rho) = \kappa \cdot (1 - \rho) \qquad (17)$$

Fig. 2. Sample images selected for our simulations.

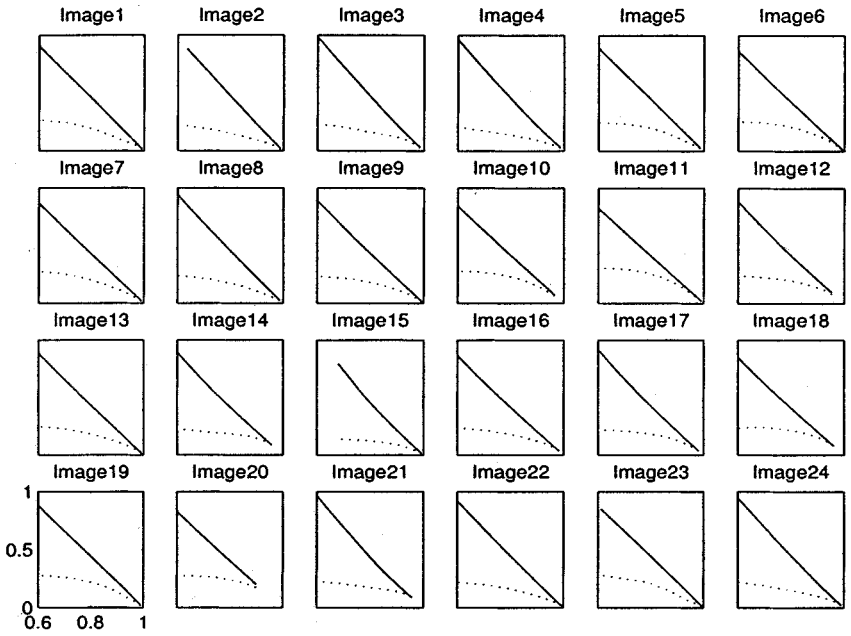

Fig. 3. Plots of $Q_{nz}(\rho)$ (solid) and $Q_z(\rho)$ (dotted) for the 24 sample images with wavelet transform and uniform threshold quantization. The x axis represents the percentage of zeros ρ while the y-axis represents the pseudo coding bit rate. All the plots have the same coordinate system.

where κ is a constant. Besides this, for each sample image, $Q_z(\rho)$ also has a rather simple behavior. In Fig. 4, we plot these two characteristic rate curves in the q-domain. It can be seen that in the q-domain, they have large image-dependent variations and highly nonlinear behaviors. The unique behaviors of $Q_{nz}(\rho)$ and $Q_z(\rho)$ exist not only for wavelet coding, but also for the DCT-based image coding. For each sample image in Fig. 2, according to their definitions, we generate $Q_{nz}(\rho)$ and $Q_z(\rho)$ with DCT and JPEG-style quantization, and plot them in Fig. 5. The above observations also hold.

2) DCT Video Coding: Next, we show that unique properties of $Q_{nz}(\rho)$ and $Q_z(\rho)$ exist not only for still images, but also for motion-compensated pictures which are the major type of source data in video coding. Let us take two QCIF video sequences "Carphone" and "News" as examples. First, we run the MPEG-2 and MPEG-4 coders on these two videos at a fixed quantization parameter 8, respectively. For each video, we output 30 sample motion-compensated difference pictures. Each sample picture is taken at every ninth frame. (The first is an I frame without motion compensation; all of the others are

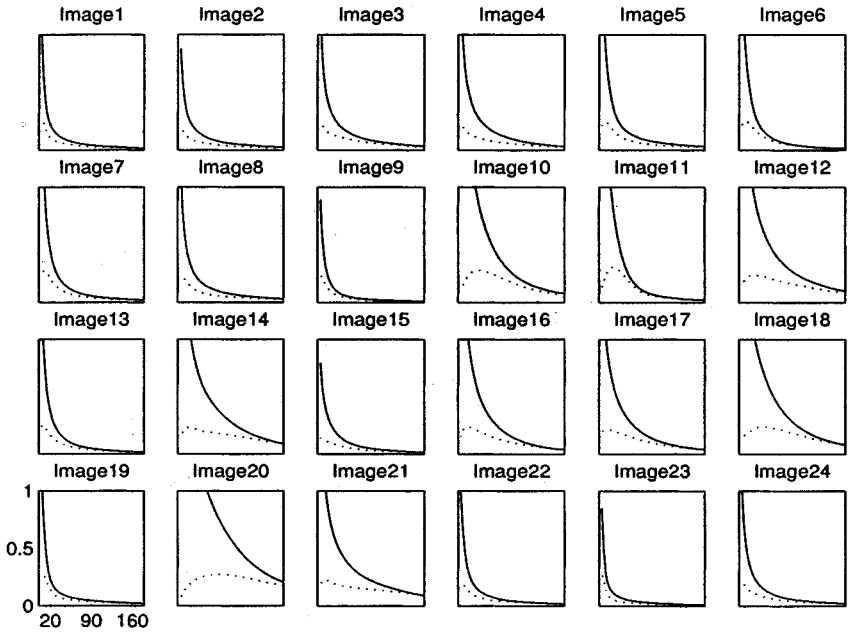

Fig. 4. Plots of $Q_{nz}(q)$ (solid) and $Q_z(q)$ (dotted) for the 24 sample images with wavelet transform and uniform threshold quantization. The x-axis represents the quantization parameter q while the y-axis represents the pseudo coding bit rate. All the plots have the same coordinate system.

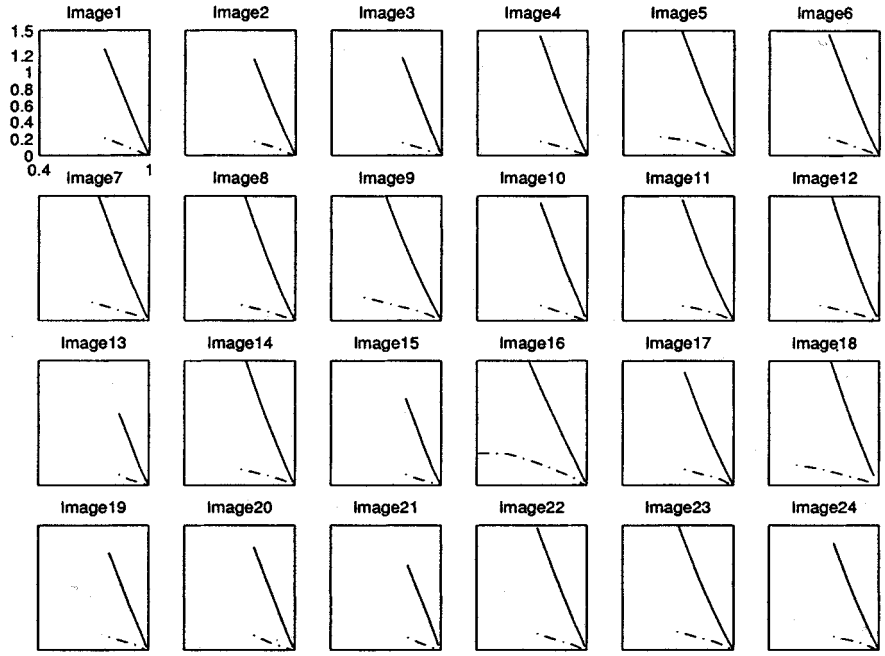

Fig. 5. Plots of $Q_{nz}(\rho)$ (solid) and $Q_z(\rho)$ (dotted) for the 24 sample images with DCT and JPEG quantization. The x axis represents the quantization parameter q, while the y axis represents the pseudo coding bit rate. All the plots have the same coordinate system.

P frames.) We plot $Q_{nz}(\rho)$ and $Q_z(\rho)$ for each sample picture from "Carphone" and "News" in Figs. 6 and 7, respectively. It can be seen that the unique properties of $Q_{nz}(\rho)$ and $Q_z(\rho)$ also exist in MPEG-2 and MPEG-4 video coding.

C. Justification of the Linearity of $Q_{nz}(\rho)$

The definition of $Q_{nz}(\rho)$ is based on the pseudo coding of the nonzero transform coefficients, which is actually the sign-magnitude representation given by (14). From the simulation results presented in the above, we observe that it has a very interesting linear behavior. In the following, we provide a theoretical justification for the linearity of $Q_{nz}(\rho)$. Note that the quantization schemes employed in the typical transform coding systems are all essentially uniform threshold quantizers. Therefore, in the following, we take the uniform threshold quantizer as an example to show the linearity of $Q_{nz}(\rho)$. It is well known that the transform coefficients have a generalized Gaussian distribution given by

$$p_{gg}(x) = \left[\frac{\nu\eta(\nu,\sigma)}{2\Gamma\left(\frac{1}{\nu}\right)}\right] \cdot e^{-[\eta(\nu,\sigma)|x|]^{\nu}} \tag{18}$$

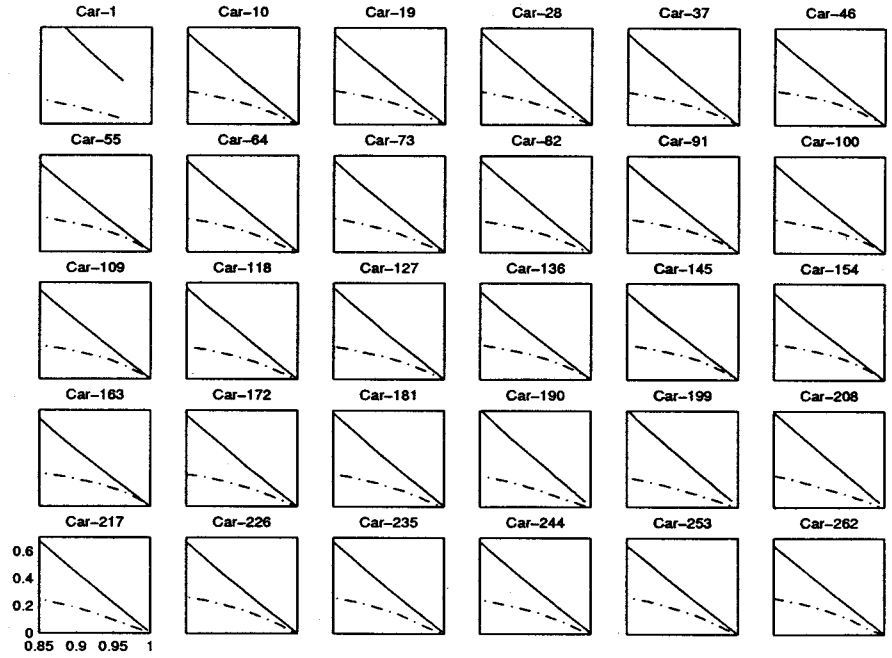

Fig. 6. Plots of $Q_{nz}(\rho)$ (solid line) and $Q_z(\rho)$ (dash–dot line) for the 30 sample difference pictures from "Carphone" coded by MPEG-2. The x axis represents the percentage of zeros ρ. All the subplots have the same coordinate system as the one at the bottom-left corner.

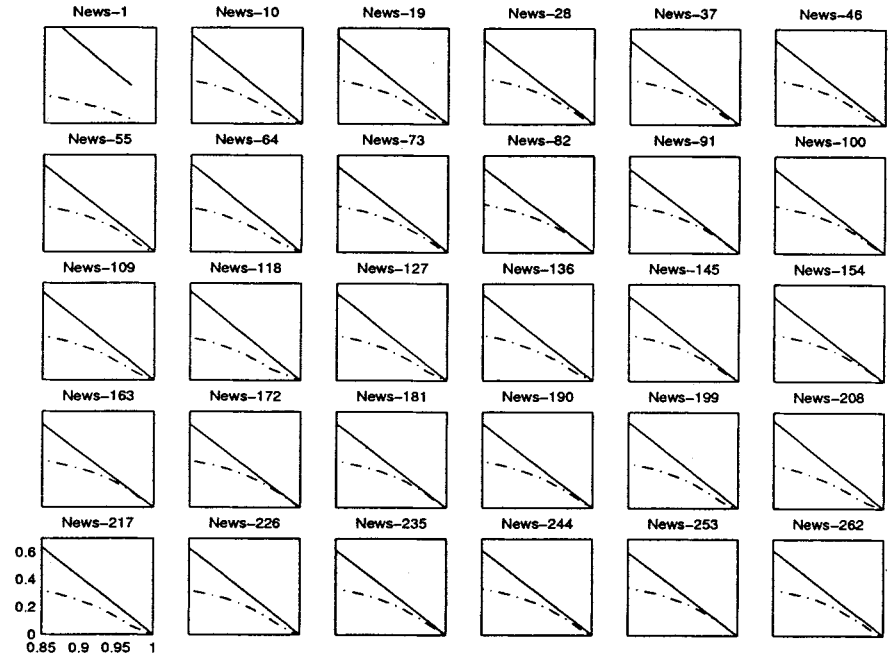

Fig. 7. Plots of $Q_{nz}(\rho)$ (solid line) and $Q_z(\rho)$ (dash–dot line) for the 30 sample difference pictures from "News" coded by MPEG-4. The x axis represents the percentage of zeros ρ. All the subplots have the same coordinate system as the one at the bottom-left corner.

where

$$\eta(\nu,\sigma)=\sigma^{-1}\left[\frac{\Gamma\left(\frac{3}{\nu}\right)}{\Gamma\left(\frac{1}{\nu}\right)}\right]^{1/2},\quad 1\le\nu\le 2. \tag{19}$$

Here, σ is the standard deviation of the transform coefficients, and ν is a model parameter which controls the shape of the distribution. For example, when $\nu=2.0$ and 1.0, $p_{gg}(x)$ becomes Gaussian and Laplacian distributions, respectively. According to (15), for any given quantization step size q, we have

$$Q_{nz}(q)=\frac{2}{M}\cdot\int_{\Delta}^{+\infty}p_{gg}(x)(\lfloor\log_2|I(x)|\rfloor+2)dx \tag{20}$$

where

$$I(x)=\left\lceil\frac{x-\Delta}{q}\right\rceil \tag{21}$$

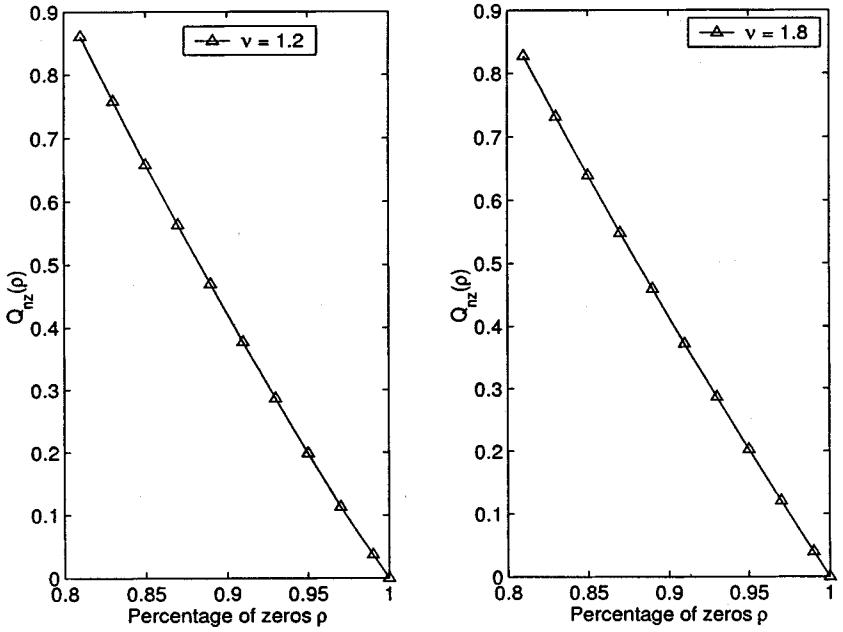

Fig. 8. Plots of the theoretically computed $Q_{nz}(\rho)$ for the generalized Gaussian distribution with different shape control parameters ν. Here, we set $\Delta = q$.

is the quantization index of x. The corresponding percentage of zeros is given by

$$\rho(q) = \frac{1}{M}\int_{-\Delta}^{+\Delta} p_{gg}(x)dx. \tag{22}$$

It is very difficult to derive a closed-form expression for $Q_{nz}(\rho)$ directly from (20) and (22). However, we can evaluate them numerically and compute a few points on $Q_{nz}(\rho)$. In Fig. 8, we plot them for different values of ν which is the shape control parameter of $p_{gg}(x)$. It can be seen that these points almost lie on a straight line. This implies that, if we assume that the transform coefficients have a generalized Gaussian distribution, $Q_{nz}(\rho)$ must be an approximately linear function.

D. Fast Estimation of $Q_{nz}(\rho)$

Since $Q_{nz}(\rho)$ is modeled as a straight line passing through the point $[1.0, 0.0]$, we need to compute only one point on it in order to estimate the whole rate function. In the following, we discuss the estimation procedure in detail for different transform coding systems.

Typical wavelet-based image coding, such SPIHT, EZW, and SR, employ the uniform threshold quantizer. After transform, we scan the subband image and generate the distribution of the transform coefficients, denoted by $\mathcal{D}(x)$. We then choose one quantization parameter q_0, and compute the corresponding value of $Q_{nz}(q_0)$ and $\rho(q_0)$ using (20) and (22) with $p_{gg}(x)$ replaced by the actual distribution $\mathcal{D}(x)$. With the two points of $[Q_{nz}(q_0), \rho(q_0)]$ and $[1.0, 0.0]$, we can construct the whole rate curve $Q_{nz}(\rho)$ with (17) where

$$\kappa = \frac{Q_{nz}(q_0)}{1 - \rho(q_0)}. \tag{23}$$

For H.263 video coding, the computation of the slope κ is outlined in [1]. In MPEG-2 coding, a perceptual quantization scheme is employed. Its quantization scheme is given by (8). As mentioned before, the perceptual quantizer is equivalent to a uniform quantizer applied to the DCT coefficients which are pre-scaled by their corresponding perceptual weights. Therefore, to compute the slope κ for the MPEG-2 video coding, we just generate the distribution information of the DCT coefficients after pre-scaling, and then apply the formula for the H.263 video coding. In MPEG-4 coding, both the H.263 style and the MPEG-2 style quantization schemes are adopted. Therefore, according to the user's configuration of the quantizer, the corresponding formula can be used to compute the value of κ.

E. Fast Estimation of $Q_z(\rho)$

In the following, by exploring the correlation between $Q_z(\rho)$ and $Q_{nz}(\rho)$, we develop a fast estimation scheme for $Q_z(\rho)$. 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 $Q_{nz}(\rho)$ is its slope κ. The feature variables for $Q_z(\rho)$ are its function values at ρ_i. Considering the characteristic rate curves plotted in Fig. 3, we choose $\rho_i = 0.70, 0.75, 0.80, 0.85, 0.90$, and 0.95. Consider the characteristic rate curves plotted in Fig. 3. For each $\rho_i, 1 \le i \le 6$, and for each of the 24 sample images, we know κ, which is the slope of $Q_{nz}(\rho)$, and $Q_z(\rho_i)$, which is the value of $Q_z(\rho)$ at ρ_i. Therefore, for each ρ_i, we have a total of 24 points of $\{[\kappa, Q_z(\rho_i)]\}$ which are depicted in Fig. 9. It can be seen that there is very strong correlation between κ and $Q_z(\rho_i)$

Fig. 10 illustrates the correlation between κ and $Q_z(\rho_i)$ for the characteristic rate curves in Fig. 5. We can see that the strong correlation also holds for JPEG coding. In our extensive simulations with a wide range of images, this strong correlation has

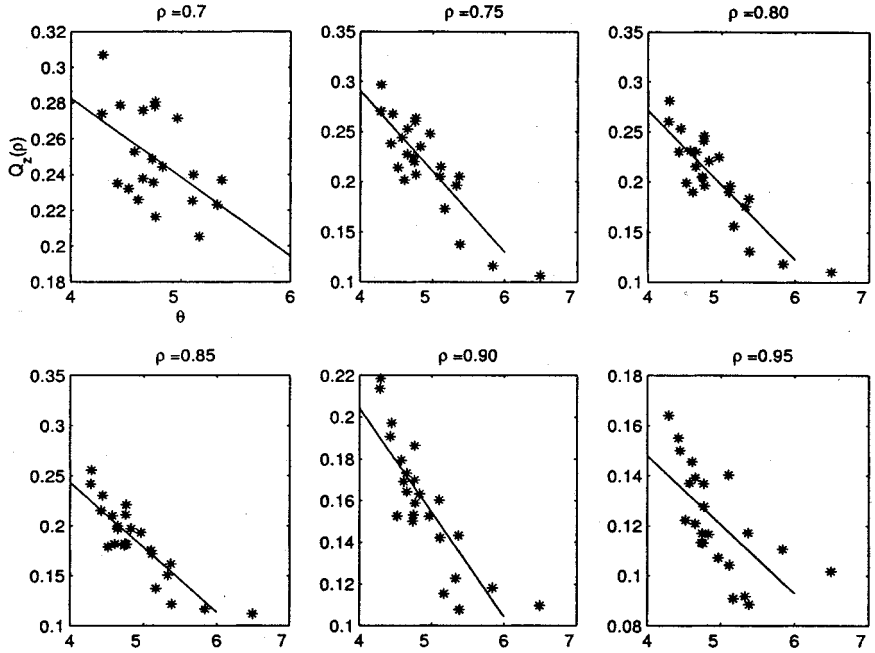

Fig. 9. Linear correlation between κ and the values $Q_z(\rho_i)$ at 0.70, 0.75, 0.80, 0.85, 0.90, and 0.95 with wavelet transform.

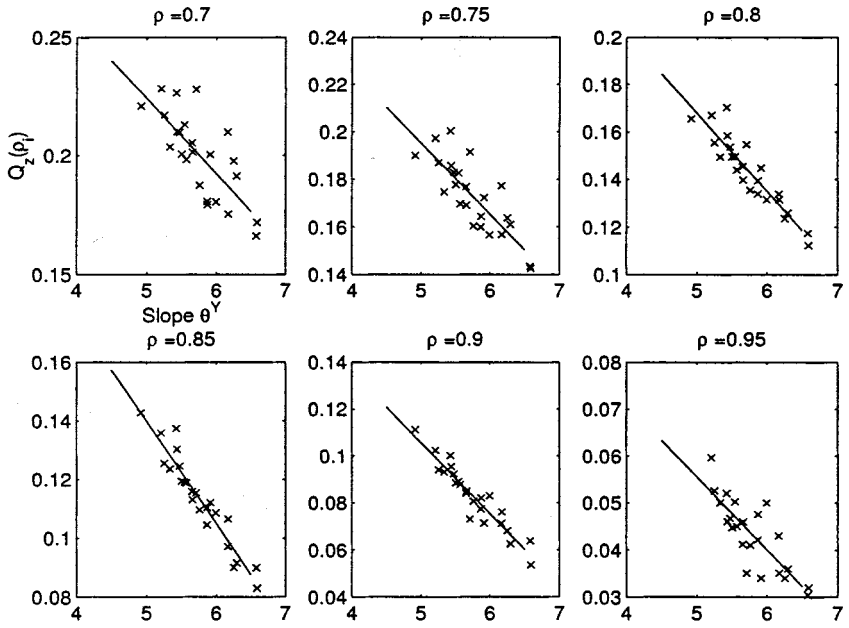

Fig. 10. Linear correlation between κ and the values $Q_z(\rho_i)$ at 0.70, 0.75, 0.80, 0.85, 0.90, and 0.95 with DCT.

been found to exist in all cases. Therefore, we have the following linear correlation model:

$$Q_z(\rho_i) = A_i\kappa + B_i \tag{24}$$

which can be employed to estimate $Q_z(\rho_i)$. The coefficients A_i and B_i are obtained by statistical regression and the corresponding linear model is also plotted in Figs. 9 and 10. Based on (24), we can estimate six points on $Q_z(\rho)$. If necessary, the whole rate curve can be constructed by linear interpolation. In [1], a cubic correlation model is employed for H.263 video coding. The correlation models for MPEG-2 and MPEG-4 video coding can also be obtained in the same way.

V. Rate Curve Decomposition

In Section IV, we have defined two rate curves $Q_{nz}(\rho)$ and $Q_z(\rho)$ to characterize the input source data, and proposed a fast algorithm to estimate them. According to our decomposition and analysis scheme, the actual rate curve in the ρ-domain $R(\rho)$ is represented by a linear combination of $Q_{nz}(\rho)$ and $Q_z(\rho)$, as shown in (13), where the coding algorithm is modeled by the decomposition coefficients $\{A(\rho), B(\rho), C(\rho)\}$. In the following, we take the JPEG coding algorithm as an example to explain how to determine the decomposition coefficients $\{A(\rho), B(\rho), C(\rho)\}$ for a specific coding algorithm.

A. Decomposition Coefficients

For the 24 sample images shown in Fig. 2, with the fast algorithms developed in Section IV, we can estimate the values of $Q_{nz}(\rho_i)$ and $Q_z(\rho_i)$ where $\rho_i = 0.65 + i \cdot 0.05, 1 \le i \le 6$. By running the JPEG coding algorithm at different quantization parameters, we can generate some points on its actual rate curve $R(\rho)$. With interpolation, we can obtain $\{R(\rho_i) | 1 \le i \le 6\}$. According to (13), we should have

$$R(\rho_i) = A(\rho_i) \cdot Q_{nz}(\rho_i) + B(\rho_i) \cdot Q_z(\rho_i) + C(\rho_i). \quad (25)$$

The values of $\{A(\rho_i), B(\rho_i), C(\rho_i)\}$ are obtained by linear regression over the sample data $\{Q_{nz}(\rho_i), Q_z(\rho_i), R(\rho_i)\}$ for the 24 sample images. The decomposition coefficients for the JPEG coding algorithm are listed in Table I. Following the same procedure, we can also obtain the decomposition coefficients for other coding algorithms, such as EZW, SPIHT, SR, MPEG-2, and MPEG-4. Once they are obtained, they are fixed during the actual rate curve estimation process. In practice, to improve the modeling accuracy, we can also update the decomposition coefficients after a certain number of frames are coded.

TABLE I
DECOMPOSITION COEFFICIENTS $A(\rho)$, $B(\rho)$, AND $C(\rho)$ AT ρ_i FOR THE JPEG CODING ALGORITHM

ρ_i	$A(\rho_i)$	$B(\rho_i)$	$C(\rho_i)$
0.70	1.2151	-0.4438	0.9005
0.75	0.8089	-0.5030	0.9201
0.80	0.6480	-0.3831	0.8449
0.85	0.5763	-0.3449	0.7856
0.90	0.5531	-0.2241	0.6808
0.95	0.4043	-0.1489	0.5845

VI. Linear Rate Regulation

Using the characteristic rate curves and the rate curve decomposition scheme introduced in Sections IV and V, we can estimate six points on the rate curve $R(\rho)$. The whole rate curve can be then obtained by linear interpolation. To further improve the estimation accuracy and robustness, we propose the following linear rate regulation scheme.

In our previous work [22], based on extensive experimental results, we have shown that in any typical image/video transform coding system, the actual coding rate function in the ρ-domain $R(\rho)$ is approximately a linear function. In other words, we have the following linear rate model

$$R(\rho) = \theta \cdot (1 - \rho). \quad (26)$$

This rate model is very accurate because it is derived from the actual coding results. The only parameter of the rate model is the slope θ. Within the context of video coding, we can estimate the value of θ from the coding statistics of previous video frame or MBs. However, in still image coding, we only have one image. In other words, there is no previous image with similar statistics available. Furthermore, typical still image coding algorithms, such as EZW, SPIHT, SR, and JPEG image coding, do not use adaptive quantization. In other words, the quantization of each transform coefficient is nonadaptive and does not depend on the quantization of its preceding coefficients. The quantization of all transform coefficients is controlled only by the picture quantization step size. The lack of the coding statistics of previous data and the nonadaptive quantization scheme make it very difficult to estimate the value of θ. Therefore, the linear rate model in (26) can not be directly used to estimate the rate function before quantization and coding. However, (26) indicates that the rate curve estimated by Sections IV and V should be a linear function. In other words, the estimated six points $\{R(\rho_i) | 1 \le i \le 6\}$ should lie on a straight line. (Because of the source modeling error, this is often not the case.) Therefore, We can utilize this linear constraint to remove the source modeling error significantly. This procedure is called *linear rate regulation*.

The linear rate regulation operates as follows. From $\{R(\rho_i) | 1 \le i \le 6\}$, we first estimate the slope θ and then construct $R(\rho)$ using the linear rate model in (26). According to the linear regression theorem, the optimum estimation of θ is given by

$$\theta = \frac{\sum_{i=1}^{6} \rho_i \sum_{i=1}^{6} R(\rho_i) - 6 \sum_{i=1}^{6} \rho_i R(\rho_i)}{6 \sum_{i=1}^{6} \rho_i^2 - \left(\sum_{i=1}^{6} \rho_i \right)^2}. \quad (27)$$

According to the linear rate model, the estimated rate curve $R(\rho)$ after linear regulation is then given by

$$R(\rho) = \frac{1}{6} \sum_{i=1}^{6} R(\rho_i) + \theta \cdot \left(\rho - \frac{1}{6} \sum_{i=1}^{6} \rho_i \right). \quad (28)$$

VII. Unified R-D Curve Estimation Algorithm

Based on the fast estimation of $Q_{nz}(\rho)$ and $Q_z(\rho)$, the rate curve decomposition, and the linear rate regulation, a unified R-D curve estimation algorithm for all typical transform-coding systems is proposed as follows.

Step 1) *Generation of the Distribution*: After applying either the DWT or the DCT transform, generate the distribution of the transform coefficients. Note that the transform coefficients are real numbers. We can approximate their distribution by the histogram of their integer parts. In the implementation of standard video coding such as MPEG and H.263, the DCT coefficients automatically have integer values. In this case, no approximation is needed. Depending on the specific quantization scheme, the distribution generation process varies. For example, in H.263, the distributions of the intracoded and intercoded MBs need to be stored separately. In JPEG and MPEG coding, we need to generate the distribution after pre-scaling of the DCT coefficients, as described in Section IV. Based on the distributions of the transform coefficients, the one-to-one mapping lookup table between q and ρ is obtained as discussed in Section II.

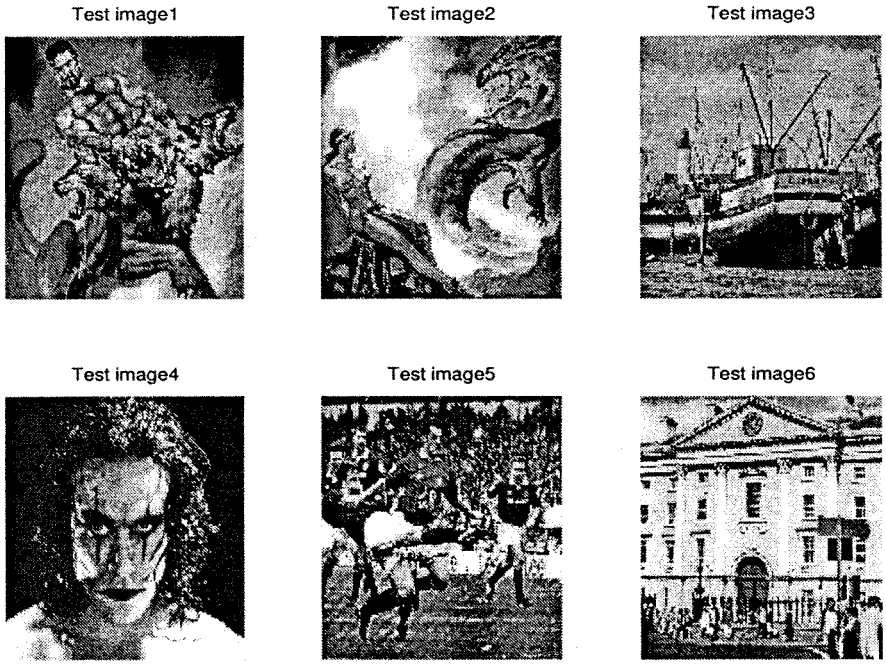

Fig. 11. Six test images for the evaluation of the proposed R-D estimation algorithm when applied to the SPIHT and SR encoders.

Step 2) *Estimate $Q_{nz}(\rho)$ and $Q_z(\rho)$*: Choose one quantization parameter q_0 and compute the corresponding $Q_{nz}(q_0)$ and $\rho(q_0)$ as discussed in Section IV-D. The slope of $Q_{nz}(\rho)$ is obtained by (23). The value of $Q_{nz}(\rho)$ at ρ_i is given by

$$Q_{nz}(\rho_i) = \kappa(1 - \rho_i). \tag{29}$$

With the linear correlation model in (24), the value of $Q_z(\rho_i)$ is determined.

Step 3) *Rate Curve Estimation*: With (25), compute $R(\rho_i)$. After linear regulation, the estimated rate curve $R(\rho)$ is given by (28). With the q-ρ mapping lookup table, $R(\rho)$ is mapped into the q-domain to obtain $R(q)$.

Step 4) *Compute Distortion Curve*: In typical image/video transform coding, each coefficient is quantized independently, the overall distortion is exactly the summation of the distortion at each coefficient. Therefore, the D-Q curve $D(q)$ can be directly computed from the distribution information.

We can see that, in the proposed algorithm, the major computation is to generate the distribution of the transform coefficients. The remaining computation involves only a few addition and multiplication operations that are carried out on the distribution. Compared to the complexity of the whole coding process, the overall complexity of the proposed estimation algorithm is very low. We can see that the R-D curves are estimated before quantization and coding.

VIII. Experimental Results

In this section, we apply the algorithm presented in Section VII to estimate the R-D curves for transform coding of still images and videos. Based on the estimated R-D functions, we can then control the output bit rate or picture quality of the encoder.

A. Application in Still Image Coding

The proposed estimation algorithm has been applied to the SPIHT and the SR encoders. We arbitrarily choose six test images as shown in Fig. 11. The estimated R-D curves and the actual ones for SR and SPIHT coding are shown in Figs. 12 and 13, respectively. It can be seen that the estimated R-D curves are very close to the actual ones curves.

The R-D estimation algorithm is also applied to JPEG coding of still images. The six test images with a wide range of R-D characteristics are shown in Fig. 14. We apply the proposed estimation algorithm to estimate their rate curves. The estimated rate curves and the actual curves are plotted in Fig. 15. The relative estimation errors of the rate curve $R(q)$ at $q_i = 0.5$, 0.8, 1.2, 2.0, 2.8, 3.2, 4.5, and 5.5 for each test image are listed in Table II. The estimation errors are very small, mostly less than 2% in their absolute values. Table III shows the relative estimation errors without linear rate regulation. Clearly, we can see that linear rate regulation significantly improves the estimation accuracy and robustness.

B. Application in DCT Video Coding

We next show that the proposed R-D estimation algorithm also works very well for DCT-based video coding. Based on the estimated R-D functions of each video frame, we can then control the output bit rate of the video encoder. We first show that, for a given MC difference video frame, the proposed algorithm can accurately estimate its R-D function. To this end, we run the MPEG-4 codec on the "Carphone" QCIF video sequence at $QP = 12$ and output the 10-th and 70-th MC difference frames. Using the algorithm presented in Section VII, we estimate their R-D curves and compare them with the actual ones generated

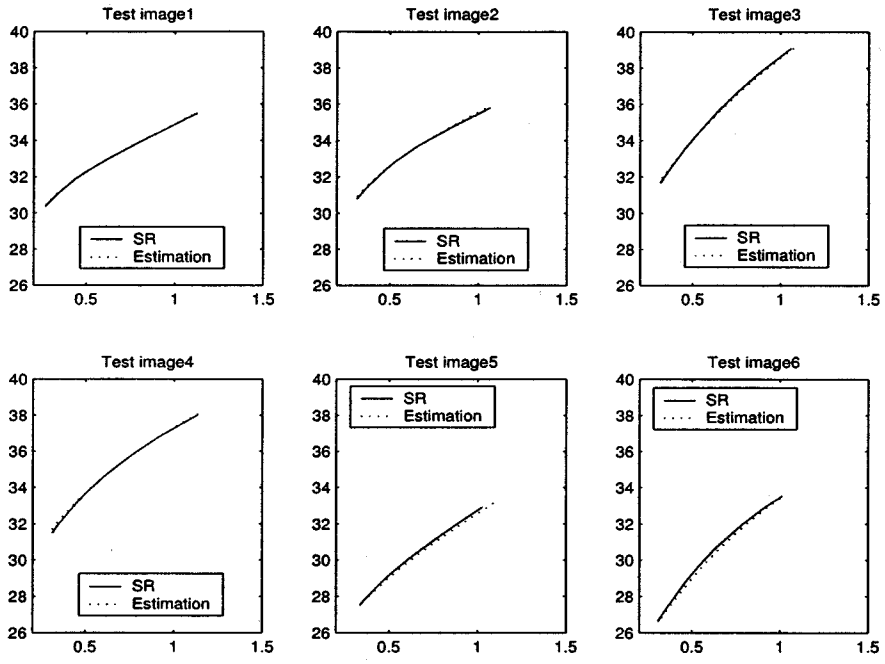

Fig. 12. R-D curve-estimation results for the SR coding system.

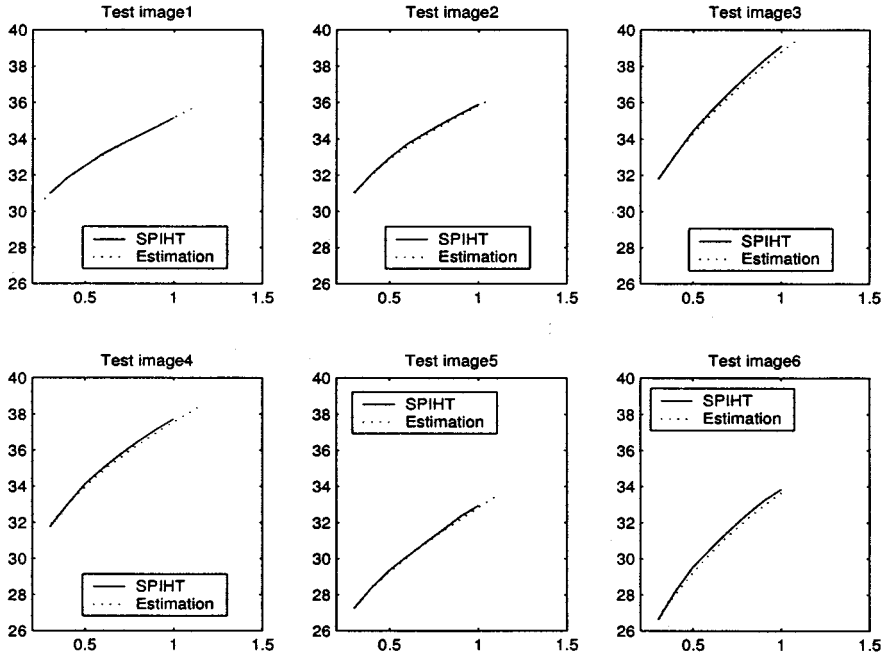

Fig. 13. R-D curve-estimation results for the SPIHT coding system.

by the MPEG-4 codec in Fig. 16. It can be seen that the R-D estimation is very accurate.

C. Frame-Level Rate Control Algorithm

With the estimated R-D functions, we can then develop a frame-level rate control algorithm for DCT video coding. The rate control process consists of the following major steps. In the first step, the target bit rate R_T of the current video frame is determined according to the channel bandwidth and buffer status, as explained in [1], [14]. In the second step, based on estimated the rate curve $R(q)$, the frame quantization parameter QP_0 can be determined to achieve the target bit rate R_T. Obviously, QP_0 satisfies $R_T = R(QP_0)$, where $R(\cdot)$ is the estimated R-Q function. We see that this rate control algorithm operates at the frame level.

It should be noted that, in standard video encoders, the quantization parameter should have an integer value between 1 and 31 [4]–[6]. However, the frame quantization parameter QP_0 obtained in the frame-level rate control algorithm is a real number. For example, QP_0 could be 5.30. If we round QP_0 to its nearest integer 5 and use it for the quantization parameter for each MB in the current frame, the actual coding bit R will be quite different from the target bit rate R_T, which is actually achieved by

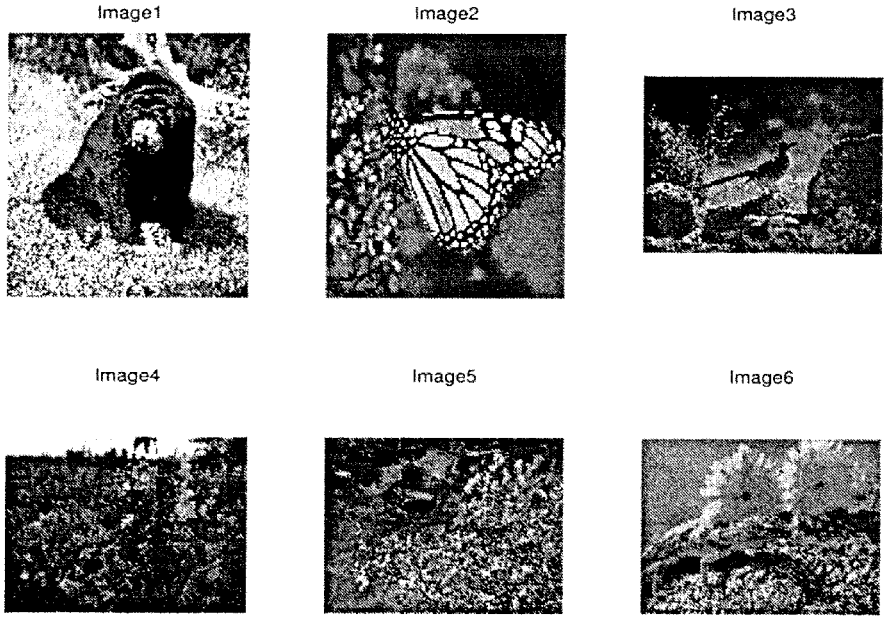

Fig. 14. The six images for testing the performance of the proposed algorithm.

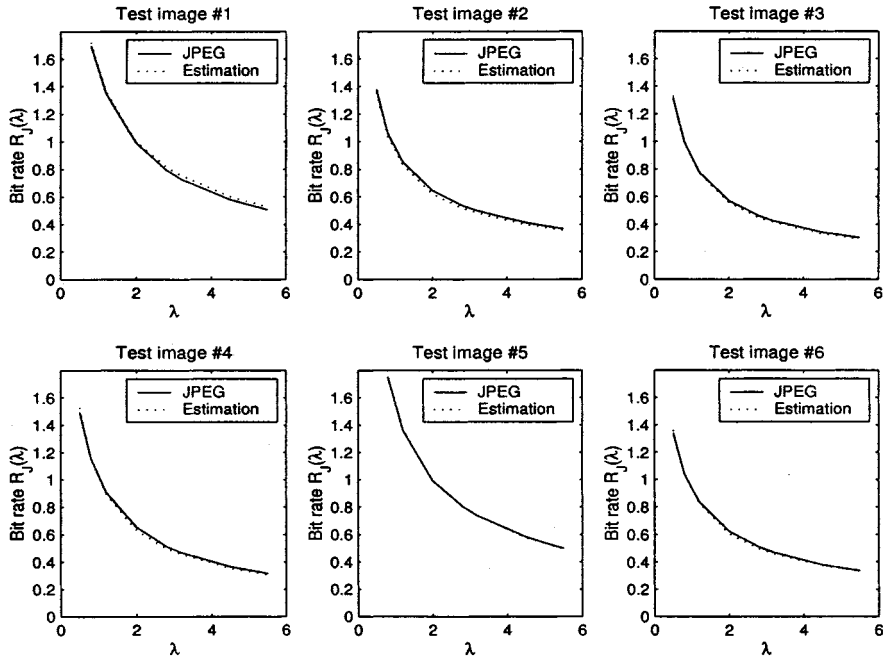

Fig. 15. The estimated rate curves and the real JPEG rate curves of the six test images. The x-axis represents the quantization parameter q.

$QP_0 = 5.30$. In the following, we propose a very simple approach to solve this problem. Let

$$QP_+ = \lceil QP_0 \rceil, \quad QP_- = \lfloor QP_0 \rfloor \tag{30}$$

which are the two closest integers to QP_0. Let $\gamma = QP_0 - QP_-$. Let ϕ be a random variable with a uniform distribution on $[0, 1]$. Each time when we determine the quantization parameter QP for a MB, we produce a sample value for ϕ. If its sample value is less than γ, we set $QP = QP_+$. Otherwise, we set $QP = QP_-$. Thus, approximately $\gamma \cdot 100$ percent of MBs in the current video frame use the quantization parameter QP_+ while all the rest use QP_-. Consequently, the average frame quantization parameter is very close to QP_0. As a result, the actual coding bit rate will be very close to the target bit rate R_T. Let S_+ and S_- be the groups of MBs which use QP_+ and QP_-, respectively. The overall activity of S_+ is almost the same as that of S_- because the MBs from each set are chosen by a uniform random variable. This is another advantage of this approach.

We implement the above rate control algorithm in MPEG-2 and MPEG-4 video coding. (The rate control algorithm for H.263 and the corresponding simulation results have been presented in [1].) Fig. 17 depicts the relative rate control errors for "Foreman" and "Coastguard" coded with MPEG-2 when the proposed algorithm and the TM5 rate control algorithm are applied. Fig. 18 shows the output coding bits of each frame in

TABLE II
RELATIVE ESTIMATION ERROR FOR JPEG CODING OF THE TEST IMAGES FIG. 14—WITH LINEAR RATE REGULATION

q_i	Rate estimation error $\mathcal{E}(q_i)$ for testing image					
	No. 1	No. 2	No. 3	No. 4	No. 5	No. 6
0.5	/	-0.1%	+0.3%	+1.6%	/	+0.6%
0.8	0.3%	-1.0%	-0.4%	-0.3%	-0.2%	-0.1%
1.2	0.2%	-1.2%	-0.2%	-0.7%	0.3%	-0.3%
2.0	0.5%	-3.0%	-1.2%	-2.1%	-0.4%	-1.1%
2.8	1.1%	-1.8%	-1.1%	-1.5%	-0.3%	-1.8%
3.2	2.8%	-0.4%	-0.9%	-2.1%	-0.1%	-1.0%
4.5	2.4%	-0.9%	-2.3%	-0.9%	-0.4%	-0.6%
5.5	2.0%	-1.2%	-0.8%	-1.1%	-0.2%	-0.1%

TABLE III
RELATIVE ESTIMATION ERROR FOR JPEG CODING OF THE TEST IMAGES FIG. 14—WITHOUT LINEAR RATE REGULATION

q_i	Rate estimation error $\mathcal{E}(q_i)$ for testing image					
	No. 1	No. 2	No. 3	No. 4	No. 5	No. 6
0.5	/	-1.4%	+1.1%	+2.3%	/	+1.4%
0.8	1.4%	-2.2%	-0.3%	-0.1%	-0.1%	-0.1%
1.2	0.8%	-2.7%	-0.9%	-1.8%	0.4%	-1.3%
2.0	1.0%	-3.9%	-2.8%	-4.0%	-0.4%	-3.1%
2.8	2.7%	-3.7%	-2.9%	-3.5%	-0.7%	-2.8%
3.2	3.8%	-3.4%	-2.1%	-2.7%	-0.1%	-2.3%
4.5	3.5%	-3.8%	-4.3%	-3.6%	-1.2%	-1.8%
5.5	4.1%	-2.8%	-2.8%	-3.1%	-0.2%	-0.2%

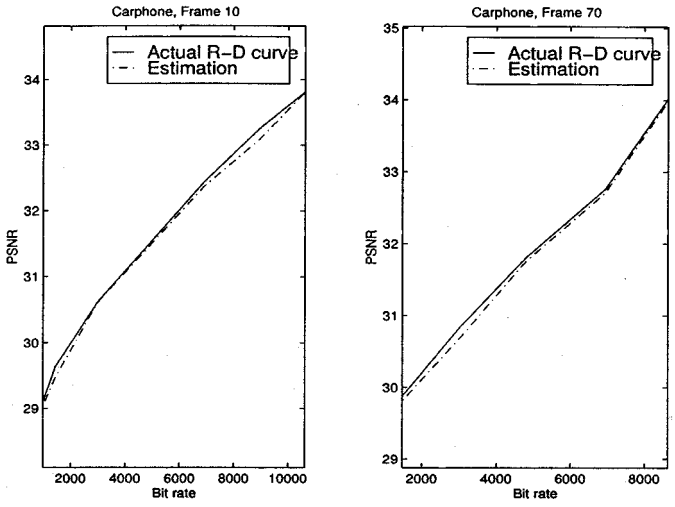

Fig. 16. R-D estimation results for Frames 10 and 70 in "Carphone" QCIF video coded by MPEG-4.

"News" QCIF video sequence coded by MPEG-4. (It should be noted that in this experiment, the whole scene is coded as one object.) It can be seen that with the proposed rate control algorithm, the actual output bit rate is more accurately matched to the bit's target.

Compared to other rate control algorithms, such as the TM5, TMN8, and VM8 algorithms, the proposed algorithm has the

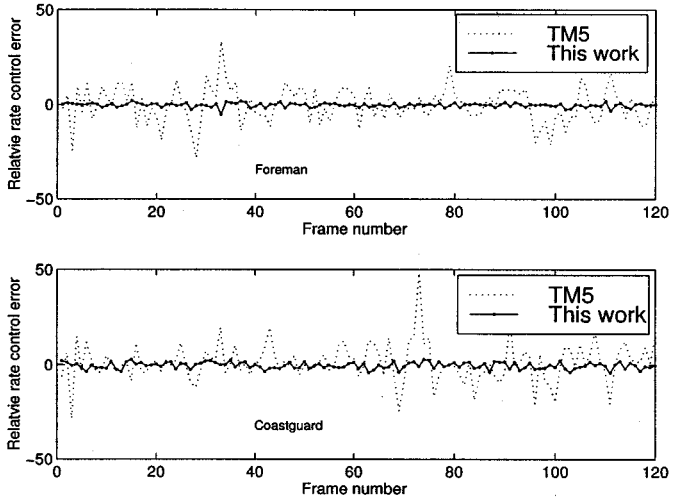

Fig. 17. The relative rate control error (in percentage) of each frame of "Foreman" and "Coastguard" coded with MPEG-2 when the proposed rate control algorithm and the TM5 algorithm are applied.

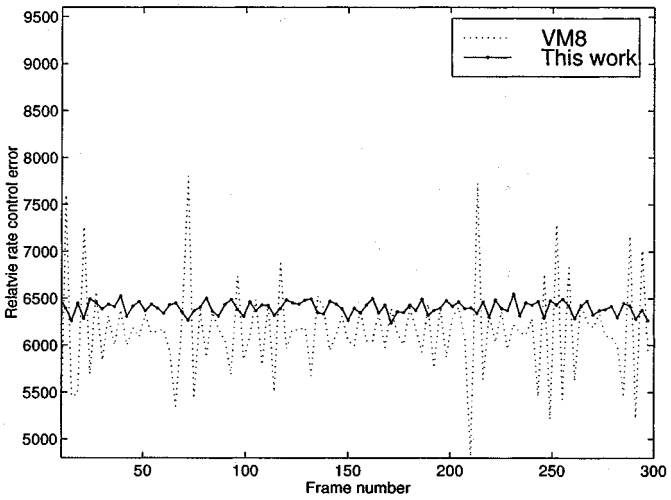

Fig. 18. The output coding bits of each frame in "News" coded with MPEG-4 when the proposed rate control algorithm and the VM8 algorithm are applied.

following advantages. First, our algorithm operates at the frame level, which implies the rate control procedure only needs to perform once per frame. However, in the MB-level TM5 and TMN8 rate control algorithms, after each MB is coded, the algorithm has to update its model parameters, which inherently increases the computational complexity. Second, based on the estimated R-D functions, the proposed algorithm select the mean quantization parameter for the current frame. In this way, the whole frame is almost uniformly quantized. However, in the TM5 and TMN8 rate control algorithms, the quantization parameter for each MB is adaptively selected based on the current coding statistics. In this way, inside one frame, the quantization settings of each MB is different (sometimes, quite different) from others. As a result, this often degrades the overall visual quality, as explained in [1]. Finally, unlike other rate control algorithms, the proposed algorithm performs the rate estimation and control for each frame independently. Therefore, it does not suffer from performance degradation at scene changes at all.

IX. Concluding Remarks

By introducing the concepts of characteristic rate curve and rate curve decomposition, a generic framework for source modeling has been developed for all typical transform coding systems. We have provided a theoretical justification for the unique property of the characteristic rate curves. A linear rate regulation scheme has also been proposed to improve the modeling accuracy and robustness. Based on this source modeling framework, a unified R-D estimation and control algorithm is proposed for image and video coding. There are two major contributions of this work. First, we present a novel methodology for R-D analysis. Compared to the conventional analytic R-D formulation, it is much more accurate. Compared to the operational R-D computation method, it has much lower complexity. The second contribution of this work is the unified R-D estimation and control algorithm, which outperforms other rate control algorithms in terms of control accuracy and robustness.

References

[1] Z. He, Y. K. Kim, and S. K. Mitra, "Low-delay rate control for DCT video coding via ρ-domain source modeling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, pp. 928–940, Aug. 2001.

[2] G. K. Wallace, "The JPEG still picture compression standard," *Commun. ACM*, vol. 34, pp. 30–44, Apr. 1991.

[3] M. W. Marcellin, M. J. Gormish, A. Bilgin, and M. P. Boliek, "An overview of JPEG-2000," in *Proc. DCC 2000*, Snowbird, UT, Mar. 2000, pp. 523–541.

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

[5] "Video Coding for Low Bit Rate Communications," ITU-T, ITU-T Recommendation H.263, version 1, 1995.

[6] T. Sikora, "The MPEG-4 video standard verification model," *IEEE Trans. Circuits Syst.Video Technol.*, vol. 7, pp. 19–31, Feb. 1997.

[7] J. M. Shapiro, "Embedded image coding using zero-trees of wavelet coefficients," *IEEE Trans. Signal Processing*, vol. 41, pp. 3445–3462, Dec. 1993.

[8] A. Said and W. A. Pearlman, "A new fast and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst.Video Technol.*, vol. 6, pp. 243–250, June 1996.

[9] M. J. Tsai, J. D. Villasenor, and F. Chen, "Stack-run image oding," *IEEE Trans. Circuits Syst.Video Technol.*, vol. 6, pp. 519–521, Oct. 1996.

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

[11] T. Berger, *Rate Distortion Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1984.

[12] H. -M. Hang and J. -J. Chen, "Source model for transform video coder and its application – part I: Fundamental theory," *IEEE Trans. Circuits Syst.Video Technol.*, vol. 7, pp. 287–298, Apr. 1997.

[13] I. H. Witten, R. M. Neal, and J. G. Cleary, "Arithmetic coding for data compression," *Commun. ACM*, vol. 30, pp. 520–540, June 1996.

[14] J. Ribas-Corbera and S. Lei, "Rate control in DCT video coding for low-delay communications," *IEEE Trans. Circuits Syst.Video Technol.*, vol. 9, pp. 172–185, Feb. 1999.

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

[16] T. Chiang and Y. -Q. Zhang, "A new rate control scheme using quadratic rate distortion model," *IEEE Trans. Circuits Syst.Video Technol.*, vol. 7, pp. 246–250, Feb. 1997.

[17] Y. Shoham and A. Gersho, "Effcient bit allocation for an arbitrary set of quantizers," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 1445–1453, Sept. 1988.

[18] K. Ramchandran and M. Vetterli, "Rate-distortion optimal fast thresholding with complete JPEG/MPEG decoder compatibility," *IEEE Trans. Image Processing*, vol. 3, pp. 700–704, Sept. 1994.

[19] *MPEG-2 Video Test Model 5*, ISO/IEC JTC1/SC29/WG11 MPEG93/457, 1993.

[20] J. Ribas-Corbera and S. Lei, "Contribution to the rate control Q2 experiment: A quantizer control tool for achieving target bit rates accurately," in *Coding of Moving Pictures and Associated Audio MPEG96/M1812 ISO/IEC JTC/SC29/WG11* Sevilla, Spain, 1997.

[21] Video Group, "Text of ISO/IEC 14 496-2 MPEG4 video VM—Version 8.0," ISO/IEC JTC1/SC29/WG11 Coding of Moving Pictures and Associated Audio MPEG 97/W1796, Stockholm, Sweden, 1997.

[22] Z. He, Y. Kim, and S. K. Mitra, "A novel linear source model and a unified rate control algorithm for H.263/MPEG-2/MPEG-4," in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing*, Salt Lake City, UT, May 2001.

[23] L.-J. Lin and A. Ortega, "Bit-rate control using piecewise approximated rate-distortion characteristics," *IEEE Trans. Circuits Syst.Video Technol.*, vol. 38, pp. 82–93, Jan. 1990.

[24] E. D. Frimout, J. Biemond, and R. L. Lagendik, "Forward rate control for MPEG recording," in *Proc. SPIE Visual Commun. Image Processing*, Cambridge, MA, Nov. 1993, pp. 184–194.

[25] A. Y. K. Yan and M. L. Liou, "Adaptive predictive rate control algorithm for MPEG videos by rate quantization method," in *Proc. Picture Coding Symposium*, Berlin, Germany, Sept. 1997, pp. 619–624.

[26] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. Image Processing*, vol. 1, pp. 205–220, Apr. 1992.

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

[28] E. Y. Lam and J. W. Goodman, "A mathematical analysis of the DCT coefficient distributions for images," *IEEE Trans. Image Processing*, vol. 9, pp. 1661–1666, Oct. 2000.

[29] MoMuSys codec, "MPEG4 verification model version 7.0 ," ISO/IEC JTC1/SC29/WG11 Coding of Moving Pictures and Associated Audio MPEG97, Bristol, U.K, 1997.

Zhihai 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.

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 1996-1999. He is currently a member of the editorial boards of *Multidimensional Systems and Signal Processing*, *Signal Processing*, *Journal of the Franklin Institute*, and *Automatika*. 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.