Optimum Bit Allocation and Accurate Rate Control for Video Coding via ρ-Domain Source Modeling

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

***Abstract*—We present a new framework for rate–distortion (R–D) analysis, where the coding rate R and distortion D are considered as functions of ρ which is the percentage of zeros among the quantized transform coefficients. In [1], we observe that, in transform coding of images and videos, the rate function $R(\rho)$ is approximately linear. Based on this linear rate model, a simple and unified rate control algorithm is proposed for all standard video coding systems, such as MPEG-2, H.263, and MPEG-4. In this paper, we further develop a distortion model and an optimum bit allocation scheme in the ρ domain. This bit allocation scheme is applied to MPEG-4 video coding to allocate the available bits among different video objects. The bits target of each object is then achieved by our ρ-domain rate control algorithm. When coupled with a macroblock classification scheme, the above bit allocation and rate control scheme can also be applied to other video coding systems, such as H.263, at the macroblock level. Our extensive experimental results show that the proposed algorithm controls the encoder bit rate very accurately and improves the video quality significantly (by up to 1.5 dB).**

***Index Terms*—Bit allocation, rate control, rate-distortion analysis, video coding.**

I. Introduction

THE ULTIMATE objective in video coding and transmission is to provide the best video quality at the receiver end, given the constraint of certain network conditions. Maximizing the picture quality implies minimizing the coding distortion of the reconstructed video. Given a bit budget, the best picture quality or minimum coding distortion can be achieved by optimum bit allocation [2] and accurate rate control [1], [3]–[6]. To be more specific, the bit allocation scheme is employed to distribute the bits budget among the video data in such a way that the overall distortion is minimized. The rate control algorithm is then employed to meet the bit's target by selecting appropriate quantization settings for the video encoder. The key issue in bit allocation and rate control is to estimate or model the rate–distortion (R–D) behavior of the video encoder. Note that the R–D behavior of an encoder is characterized by its rate–quantization (R–Q) and distortion-quantization (D–Q) functions, which are collectively called R–D functions in this work.

Manuscript received June 5, 2001; revised November 1, 2001. This paper was recommended by Associate Editor H. Sun.

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).

Digital Object Identifier 10.1109/TCSVT.2002.804883

In the classic R–D analysis [3], [7], [8], the coding bit rate is approximated by the zero-order entropy of the quantized coefficients. However, in transform coding of images and videos, especially at very low bit rates, there is a large mismatch between the zero-order entropy and the actual coding bit rate [3]. This is because, after discrete wavelet transform (DWT) or discrete cosine transform (DCT), there is still significant correlation left among the transform coefficients which can be further explored by the coding algorithm. However, the analytic entropy formulation does not take the efficient coding algorithm into account.

Since it is difficult to develop a closed-form expression to directly model the R–D behavior, an empirical approach is often used in R–D modeling and rate control. Several R–D estimation, and control algorithms have been developed within the context of video coding [3]–[6]. Some of them have been adopted as international standards; typical of these are the MPEG-2 Test Model Version 5 (TM5) rate control algorithm [11], the H.263 Test Model Near-term Version 8 (TMN8) algorithm [5], and the MPEG-4 Verification Model Version 8 (VM8) algorithm [4]. They are widely used in practical coding applications. Besides these standard rate control algorithms, many other algorithms have been proposed to target different applications. A parametric R–D model has been proposed by Tao *et al.* [14] for frame-level MPEG video coding. An approach based on a normalized parametric R–D model [13] has been developed for H.263-compatible video codecs.

Among the rate models used in these rate control algorithms, some are based on the modified version of the classical R–D functions which lead to logarithmic expressions [3], [5], [14]. Mathematical expressions of other types, such as power [6], spline [12], and polynomial [4], have also been employed. It can be seen that these R–D models have complex and highly nonlinear expressions. However, they still often suffer from relatively large control errors and performance degradation at scene changes. In addition, each of these R–D models and control algorithms targets at a specific video coding system.

By introducing a new framework for R–D analysis, called ρ-domain R–D analysis, we developed a linear rate model in [1]. Based on this simple rate model, a unified ρ-domain rate control (ρ-RC) algorithm is proposed for all the standard video coding systems, such as MPEG-2, H.263 [15], [32] and MPEG-4 [16]. Compared to other algorithms reported in the literature, the ρ-RC algorithm controls the encoder bit rate much more accurately and robustly [1].

In this paper, based on the proposed R–D analysis framework, we develop a distortion model in the ρ domain. Coupled with the linear rate model, a ρ-domain optimum bit allocation scheme is developed. This bit allocation scheme is applied to MPEG-4 video coding to distribute the bits among the video objects in the scene. The bit budget allocated to each video object is achieved by the ρ-RC rate control algorithm. For nonobject-based video coding, such as H.263, we propose a macroblock (MB) classification scheme, and apply the proposed algorithm to the macroblock classes. Our experimental results show that with the proposed optimum bit allocation and accurate rate control scheme, the coded picture quality is significantly improved (by up to 1.5 dB).

The rest of this paper is organized as follows, In Section II, we introduce the ρ-domain R–D analysis framework. For the integrity of this paper, a brief review of the linear rate model and ρ-RC algorithm developed in [1] is provided in Section III. The ρ-domain distortion model is developed in Section IV. Section V presents the ρ-domain bit allocation scheme. In Section VI, we apply the proposed bit allocation and rate control algorithm to MPEG-4 object-based video coding. In Section VII, with macroblock classification, the algorithm is applied to H.263 video coding. The respective experimental results are included in each section. Section VIII provides some concluding remarks.

II. ρ-Domain R–D Analysis

It has been observed that zeros play a key role in transform coding, especially at low bit rates [17]. All typical coding algorithms treat zeros in a special way and address most of the effort to efficient coding of zeros. 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 [9], [10]. 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 [15], [32]. 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 ρ and q. 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*.

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 typical transform coding systems, such as JPEG, MPEG-2, H.263, and MPEG-4, each transform coefficient is quantized separately. In the following, we take the H.263 coding as an example to explain how to compute the one-to-one mapping between q and ρ. In the H.263-style quantization scheme, the quantization index of a DCT coefficient x is given by

$$I[x;\, q] = \begin{cases} \text{Round}\left(\dfrac{x}{8}\right), & \text{if } x \text{ is a DC coefficient in an intra-MB} \\ UTQ(2q,\, 2q;\, x), & \text{if } x \text{ is an AC coefficient in an intra-MB} \\ UTQ(2q,\, 2.5q;\, x), & \text{if } x \text{ is a coefficient in an inter-MB} \end{cases} \tag{1}$$

where “UTQ” represents the uniform threshold quantization

$$UTQ[q,\, \Delta;\, x] = \begin{cases} 0, & \text{if } |x| \le \Delta \\ \left\lceil \dfrac{x-\Delta}{q} \right\rceil, & \text{if } x > +\Delta \\ \left\lfloor \dfrac{x+\Delta}{q} \right\rfloor, & \text{if } x < -\Delta. \end{cases} \tag{2}$$

Here, Δ is the dead zone threshold. Let $\mathcal{D}_0(x)$ and $\mathcal{D}_1(x)$ be the distributions of the DCT coefficients in the intracoded and intercoded macroblocks, respectively. Note that, in general, the DC coefficients from the intracoded macroblocks 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:

$$\rho(q) = \frac{1}{L}\int_{-2q}^{+2q} \mathcal{D}_0(x)\,dx + \frac{1}{L}\int_{-2.5q}^{+2.5q} \mathcal{D}_1(x)\,dx \tag{3}$$

where L is the number of coefficients in the current video frame. Note that in the H.263 codec, the DCT coefficients are rounded to integers [18]. Therefore, $\mathcal{D}_0(x)$ and $\mathcal{D}_1(x)$ are actually histograms of the DCT coefficients, and (3) becomes

$$\rho = \frac{1}{L}\sum_{|x|<2q} \mathcal{D}_0(x) + \frac{1}{L}\sum_{|x|<2.5q} \mathcal{D}_1(x). \tag{4}$$

It can be seen that (4) only involves a few addition operations.

III. ρ-Domain Linear Rate Model and Rate Control

For the integrity of this paper, in this section, we provide a brief review of the ρ-domain linear rate model and rate control algorithm presented in [1].

A. Linear Rate Model

Based on extensive experimental results, we have shown in [1] that, in all standard video coding systems such as MPEG-2, H.263, and MPEG-4, for different types of source data such as I, P, and B frames, the base layer and enhancement layer, the rate function in the ρ-domain is approximately a linear function. In other words, $R(\rho)$ has the following expression:

$$R(\rho) = \theta \cdot (1-\rho) \tag{5}$$

where θ is a constant. In the following experiment, we show that this linear rate model is also valid in object-based MPEG-4 coding. We run the MPEG-4 codec on the “News” sequence (with two objects: foreground and background objects) at different quantization parameters, and generate several points $\{\rho_i,\, R_i\}$ on the rate curve $R(\rho)$. Let $\mathcal{C}(\rho_i,\, R_i)$ be the correlation between $\{\rho_i\}$ and $\{R_i\}$. In Fig. 1, we plot the value of $-\mathcal{C}(\rho_i,\, R_i)$ for each coded video object plane (VOP). It can be seen that $\mathcal{C}(\rho_i,\, R_i)$ is very close to -1, which implies that there is a very strong linear relationship between $\{\rho_i\}$ and $\{R_i\}$. The rate model in (5) is very simple due to its linear expression. It is also very accurate because it represents the actual coding bit rate of the image/video encoder.

In Table I, we list the average percentage of zeros among the quantized DCT coefficients for a wide range of coding bit

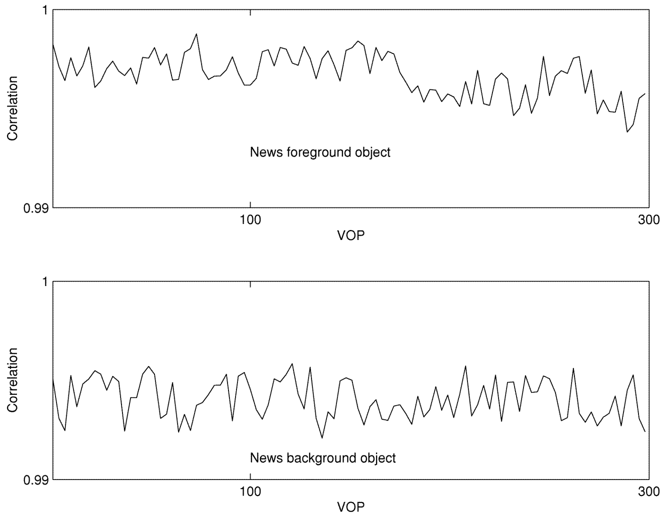


Fig. 1. Correlation coefficient (inverse) between the percentage of zeros ρ and the coding bit rate R for each VOP in the foreground and background objects of "News."

rates. The coding algorithm is MPEG-4. The four test videos are "Carphone," "Akiyo," "Foreman," and "Football" in QCIF format coded at 15 frames per second (fps). The last row of the table lists the corresponding peak signal-to-noise ratio (PSNR) for 960 kbps. We can see that these bit rates and PSNR values are much higher than those required in practical video coding applications. Even for 960 kbps, which is an extremely high coding bit rate for QCIF videos, the average value of ρ is still above 70%. Therefore, for practical purposes, we only need to consider ρ larger than 70%. For the coding bit rates commonly used in practical applications, such as those less than 384 kbps, ρ is mostly larger than 90%.

B. Estimation of θ

The only parameter of the linear rate model in (5) is the slope θ. In [1], we propose an adaptive estimation algorithm to estimate the value of θ for each frame. Let N_c be the number of the coded macroblocks in the current frame. Note that in a 16 × 16 macroblock, there are totally 384 luminance and chrominance coefficients. Let R_c be the number of bits already used to encode these N_c macroblocks. Let ρ_c be the number of zeros produced by the quantization of these macroblocks. According to (5), θ can be adaptively estimated as follows:

$$\theta = \frac{R_c}{384 \cdot N_c - \rho_c}. \tag{6}$$

The estimated θ is then applied to rate control of the current macroblock.

C. ρ-Domain Rate Control

Based on the linear rate model and the adaptive estimation of θ, a ρ-domain rate control (ρ-RC) algorithm is proposed. Let the target bits per frame (or channel bandwidth) be W. Let the encoder buffer size be B_T and the number of bits in the buffer be B_0. The available bits for coding the current frame is

$$R_T = W - B_0 + \alpha \cdot B_T \tag{7}$$

TABLE I
AVERAGE PERCENTAGE OF ZEROS FOR A WIDE RANGE OF CODING BIT RATES

Rate	Car	Akiyo	Foreman	Football
64 Kbps	98.78%	98.49%	98.85%	98.33%
96 Kbps	97.95%	97.59%	98.05%	98.02%
128 Kbps	97.08%	96.67%	97.21%	97.17%
256 Kbps	93.58%	92.77%	93.68%	93.63%
384 Kbps	90.06%	89.21%	90.05%	90.13%
512 Kbps	86.38%	85.84%	86.57%	86.77%
960 Kbps	73.22%	83.16%	73.38%	75.68%
PSNR	45.04 dB	48.02 dB	44.23 dB	37.75 dB

where the target buffer level α is, by default, set to be 0.2 [5]. Let M be the number of macroblocks in a video frame. (For QCIF videos, M is 99.) The quantization parameter is determined by the following steps.

Step 1 *Initialization*: Before encoding the first macroblock, set $R_c = \rho_c = 0$. Generate the distributions $\mathcal{D}_0(x)$ and $\mathcal{D}_1(x)$ for the DCT coefficients in the intra and inter macroblocks, respectively. Set $\theta = 7$, which is its average value.

Step 2 *Determine the quantization parameter* q: Suppose the current MB number is N_c. The number of coefficients in the remaining uncoded MBs is $384 \cdot (M - N_c)$. Note that there are still $R_T - R_c$ bits available. According to (5), the percentage of zeros to be produced by the quantization of the remaining macroblocks should be

$$\rho = 1 - \frac{1}{\theta} \cdot \frac{R_T - R_c}{384 \cdot (M - N_c)}. \tag{8}$$

Based on the one-to-one mapping between ρ and q, the step size q is then determined. The current macroblock is quantized with q and entropy coded.

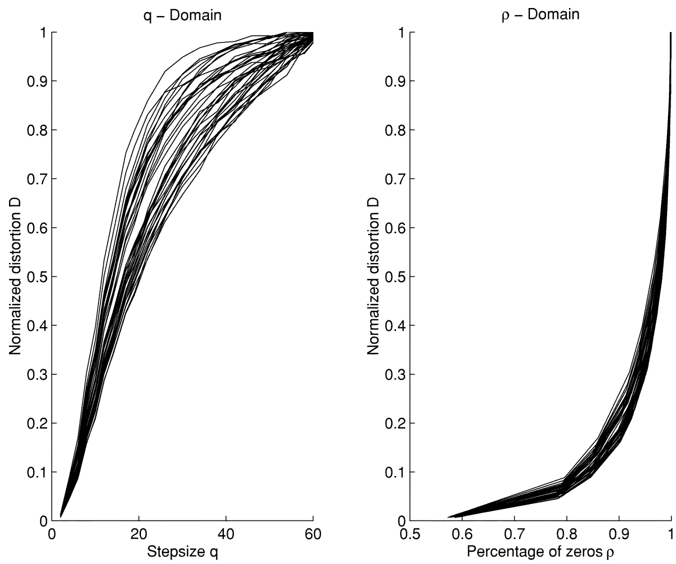
Step 3 *Update*: Let ρ_0 and R_0 be the number of zeros and number of bits produced by the current macroblock, respectively. Set $\rho_c = \rho_c + \rho_0$, $R_c = R_c + R_0$, and $N_c = N_c + 1$. If $N_c \geq 10$, update the value of θ according to (6). At the same time, subtract the frequencies of the DCT coefficients in the current macroblock from $\mathcal{D}_0(x)$ if it is an intra macroblock, or from $\mathcal{D}_1(x)$ if it is an inter macroblock.

Step 4 *Loop*: Repeat steps 2 and 3 for the next macroblock until all the macroblocks in the current frame are encoded.

It can be seen that the above rate control algorithm is conceptually simple with very low computational complexity. It only involves addition and few simple multiplication operations. Therefore, the proposed algorithm also has a low implementation cost. The experimental results presented in [1] show that our ρ-RC algorithm outperforms other rate control algorithm reported in the literature by providing much more accurate and robust rate control.

IV. ρ-DOMAIN DISTORTION MODEL

In the classical R–D analysis, the distortion D (the mean square error between the reconstruction frame and the original



(a) (b)

Fig. 2. Distortion curves of each frame in Foreman.qcif in (a) the q domain and (b) the ρ domain.

one) is considered to be a function of the quantization parameter q, denoted by $D(q)$. From [1], we know that the rate function $R(\rho)$ has a unique behavior in the ρ domain. This motivates us to also study the distortion function in the ρ domain. Let $\overline{D} = D/\sigma^2$ be the normalized distortion, where σ^2 is the picture variance. In Fig. 2, we plot the normalized distortion function in the q domain $\overline{D}(q)$ and the one in the ρ-domain $\overline{D}(\rho)$ for each video frame of "Foreman" coded by H.263. Two observations are made from these plots. First, in the q domain, $\overline{D}(q)$ is defined over an infinite range $[0, +\infty)$. This is because theoretically the quantization step size can be arbitrarily large. However, in the ρ domain, the distortion function $\overline{D}(\rho)$ is defined within a finite range [0, 1], since the largest value of ρ is 100%. Mathematically, it is more convenient to analyze functions over a finite range than an infinite range. Note that when $\rho \to 0$, which means that the quantization parameter q is very small, we have $\overline{D}(\rho) \to 0$. When $\rho \to 1$, which means that the quantization parameter q is very large, we have $\overline{D}(\rho) \to 1$. Second, in the q-domain, for different video frames, the plots of $\overline{D}(q)$ are quite different from each other. However, in the ρ domain, for different video frames, the variation of $\overline{D}(\rho)$ is very small. This implies that the distortion function has a more robust and regulated behavior in the ρ domain than in the q domain. We observe that, for each video frame, $D(\rho)$ has an exponential behavior and can be well approximated by

$$D(\rho) = \sigma^2 e^{-\alpha(1-\rho)} \tag{9}$$

where α is a constant that normally ranges from 10 to 20. Assume that the DCT coefficients have a Laplacian distribution [19] given by

$$p_l(x) = \frac{\lambda}{2} e^{-\lambda|x|}. \tag{10}$$

As we know, the relationship between λ and σ^2 is given by

$$\sigma^2 = \frac{2}{\lambda^2}. \tag{11}$$

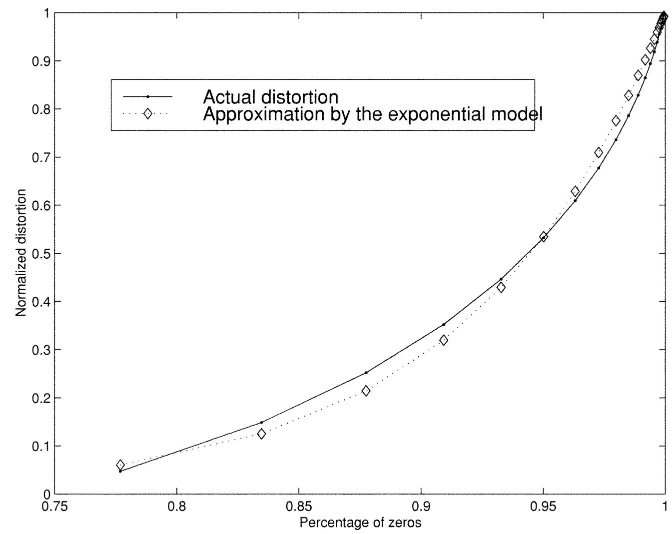


Fig. 3. Comparison between the actual ρ-domain distortion function and the exponential distortion model.

For a uniform threshold quantizer with a dead-zone threshold $\Delta = (0.5 + b) \cdot q$, where b is a nonnegative constant, the quantization distortion is given by

$$D(q) = 2\int_0^{\Delta} p_l(x)x^2\,dx + 2\sum_{i=0}^{\infty} \cdot \int_{iq+\Delta}^{(i+1)q+\Delta} p_l(x)\left|x - \left(i + \frac{1}{2}\right)q - \Delta\right|^2 dx. \tag{12}$$

The corresponding percentage of zeros is given by

$$\rho = \int_{-\Delta}^{\Delta} \frac{\lambda}{2} e^{-\lambda|x|}\,dx \tag{13}$$

$$= 1 - e^{-\lambda\Delta}. \tag{14}$$

Combining (12) and (14), we have

$$D(\rho) = \frac{1 + \eta^a - 2\eta + (1-a)(1+\eta^a)\eta\ln\eta}{\lambda(1+\eta^a)} \tag{15}$$

where $a = 1/(1+2b)$ and $\eta = 1 - \rho$. Obviously, this distortion function is highly nonlinear and complex. It will be very difficult to develop a close-form formula for the optimum bit allocation based on such a complex distortion model. However, we observe that $D(\rho)$ given by (15) can be closely approximated by (9), which is much simpler. To show this, in Fig. 3, we plot the functions in (15) and (9) for $\lambda = 0.35$ and $b = 0.5$. We can see that the simple exponential distortion model approximates the actual distortion function very closely. For this reason, we use (9) instead of (15) as our ρ-domain distortion model.

V. Optimum Bit-Allocation Scheme

Based on the distortion model in (9) and the linear rate model in (5), a ρ-domain optimum bit allocation scheme is developed in this section.

A. Bit Allocation in Brief Review

Before the formulation of our bit allocation scheme, we briefly review some existing bit allocation schemes developed in the literature. In transform coding of images and videos, bit allocation is employed to distribute the bit's budget among different groups of transform coefficients to achieve the minimum overall quantization distortion. The problem of optimum bit allocation was first addressed by Huang and Schultheiss [20], where only an approximate solution to the problem was provided. Further improvements have been suggested in [2], [21], and [22] within the context of source quantization and coding. The optimum bit allocation scheme can be applied to various image-coding algorithms to improve their coding performance, such as JPEG [23] and wavelet-based image coding [24]. In video coding, bit allocation can be incorporated into the rate control algorithm to further extend the capability of the control algorithm and to improve the video presentation quality [5], [11], [25].

The optimum bit allocation is carried out based on the R–D functions of the encoder. The analytic formulas of the R–D functions are used to derive the closed-form expression for the optimum bit allocation scheme as in [2], [5], [20]–[22]. As mentioned in Section I, the analytic models in the conventional R–D analysis often suffer from relatively large estimation error. As a result, the optimum bit allocation based on these R–D models cannot be truly optimum [6]. For this reason, in practical image/video coding, more accurate operational R–D curves are employed to perform the optimum bit allocation [12], [24]. Since the generation of the operational R–D curves often has very high computational complexity, this type of operational bit allocation scheme does not work efficiently in practical video applications, especially in real-time video coding and transmission. In the previous sections, we have shown that the R–D functions have unique properties in the ρ domain and developed simple and accurate models for the R–D functions. Based on these models, we can then develop an optimum bit allocation scheme in the ρ domain.

B. ρ-Domain Bit Allocation

In transform coding of images and videos, we need to take two major steps to achieve the best picture quality. The first step is the optimum bit allocation. Specifically, we need to determine the number of bits assigned to each data source in such a way that the overall distortion is minimized. Here, "source" is a generic term. In video coding, it could be a frame, a video object, or a group of macroblocks inside one frame. In the second step, we need to accurately select the quantization parameter to meet the bit budget for each source, which is exactly the problem of rate control. We have solved this problem by developing the ρ-RC rate control algorithm. So, the only remaining issue is to develop a ρ-domain optimum bit allocation scheme.

In the ρ domain, the rate and distortion functions for each input source are given by (5) and (9), respectively. Let $\{S_i | 1 \leq i \leq K\}$ be the input sources. For each S_i, we have

$$R_i(\rho_i) = \theta_i \cdot (1-\rho_i) \tag{16}$$
$$D_i(\rho_i) = \sigma_i^2 e^{-\alpha_i(1-\rho_i)}. \tag{17}$$

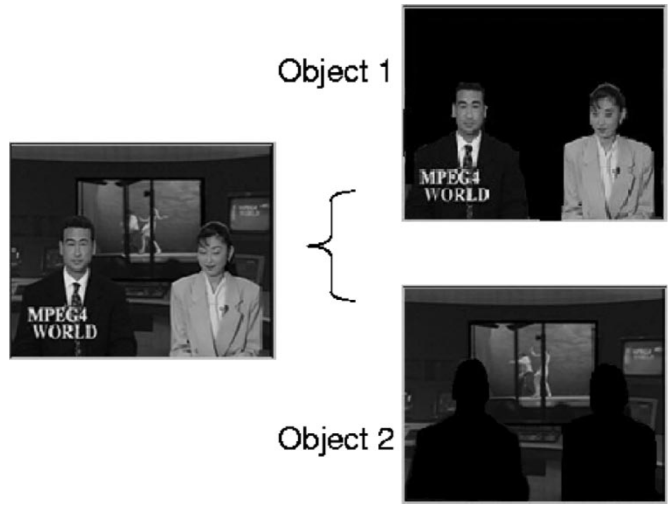


Fig. 4. Illustration of the video objects segmentation in MPEG-4 coding.

The optimum bit allocation problem can then be formulated as follows:

$$\min_{\rho_i} \quad \sum_{i=1}^{K} \sigma_i^2 e^{-\alpha_i(1-\rho_i)} \cdot N_i \tag{18}$$
$$\text{s.t.} \quad \sum_{i=1}^{K} \theta_i \cdot (1-\rho_i) \cdot N_i = R_T \tag{19}$$

where N_i is the size of S_i and R_T is total number of bits available. With the Langrange multiplier, the constrained minimization problem can be converted to the following unconstrained problem:

$$\min_{\rho_i} \quad \sum_{i=1}^{K} \sigma_i^2 e^{-\alpha_i(1-\rho_i)} \cdot N_i + \lambda \cdot \left[\sum_{i=1}^{K} \theta_i \cdot (1-\rho_i) \cdot N_i - R_T\right]. \tag{20}$$

Following the same minimization procedure as described in [27], we obtain the optimum number of bits for each input source

$$R_i = \xi_i N_i \ln \frac{\sigma_i^2}{\xi_i} + \frac{\xi_i N_i}{\sum_{i=1}^{K} \xi_i N_i} \left(R_T - \sum_{i=1}^{K} \xi_i N_i \ln \frac{\sigma_i^2}{\xi_i}\right) \tag{21}$$

where $\xi_i = \theta_i / \alpha_i$. In the following, we apply this optimum bit allocation scheme to practical video coding.

VI. Scalable Rate Control for MPEG-4

The ISO MPEG-4 video coding supports content-based interactivity which allows the access and manipulation of video objects in the compression domain [16]. To this end, each video frame is segmented into several objects associated with some physical meaning, such as foreground people and background scene, as illustrated in Fig. 4. Each video object is then coded separately. The MPEG-4 output bit stream syntax also allows the separate decoding and reconstruction of each video object.

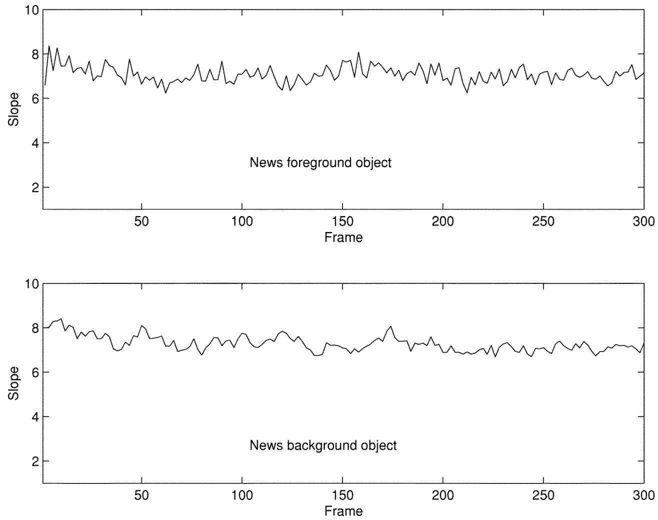


Fig. 5. Value of slope θ for each VOP in "News" sequence.

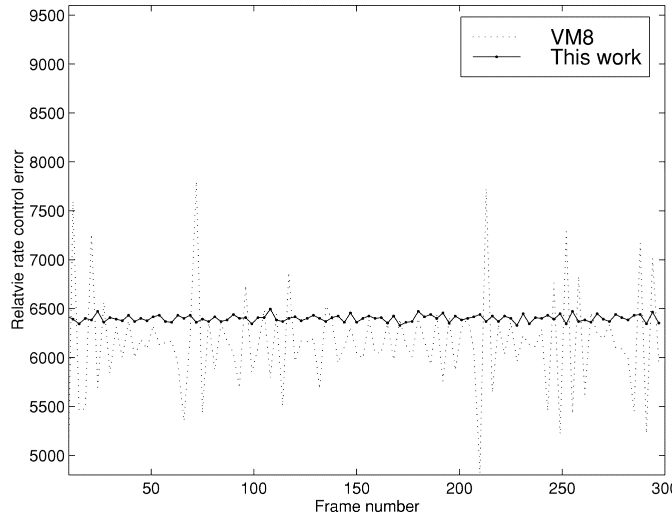


Fig. 6. Bits per frame when (solid line) the proposed algorithm and (dotted line) the VM8 algorithm are applied to the MPEG-4 codec.

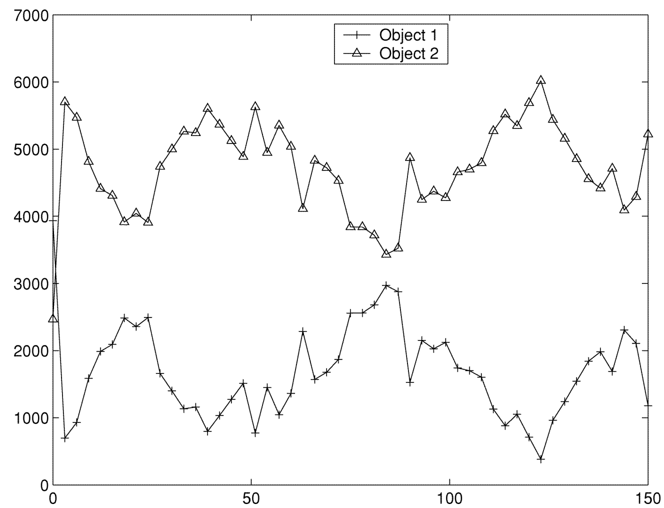


Fig. 7. Bits assigned to each video object in the "News" sequence when the weighted optimum bit allocation scheme is applied to the MPEG-4 codec. The x axis represents the frame number; while the y axis represents the number of bits per frame.

In this way, scene editing, such as adding, deleting, or moving a video object can be performed directly on the bit stream. With this type of content-based coded representation of video data, the video information can used and presented in a much more flexible way.

In the rate control for MPEG-4 coding, the available bit budget R_T for each video frame can be determined from the channel bandwidth and the buffer status. The optimum bit allocation scheme developed in Section V can then be employed to efficiently distribute these bits among the video objects to maximize the overall picture quality. Once the bits' target R_i for each video object is obtained, the macroblock-level ρ-RC rate control algorithm, proposed in [1], can then be employed to control the encoder to achieve the bits target R_i.

A. Model Parameters

Note that, in the bit allocation scheme just described, there are two model parameters θ_i and α_i to be determined. In practical video coding, the parameter θ of the current object can be determined by the coding statistics of the same object in the previous frame of the same type. To be more specific, after coding the kth frame, we already know the number of bits R_i used for coding the object VO_i, the percentage of zeros ρ_i produced by VO_i. With (16), the values of θ_i can be determined as follows:

$$\theta_i = \frac{R_i}{1-\rho_i}. \tag{22}$$

It is then used in the bit allocation for object VO_i in the current $(k+1)$th video frame. Note that, after scene segmentation, each video object becomes more homogeneous. The temporal variation of the model parameter θ is significantly reduced. To show this, we plot θ for each VOP in "News" in Fig. 5. It can be seen that the variation of θ is very small. For this reason, the proposed frame-level estimation of θ works quite well in practice. For the model parameter α_i, there is a direct and even more accurate way to estimate it. Using (12) and (14), we can compute D_i and ρ_i for a given quantization parameter q. According to the distortion model (17), α_i can be determined as follows:

$$\alpha_i = \frac{1}{1-\rho_i} \ln \frac{\sigma_i^2}{D_i}. \tag{23}$$

B. Experimental Results

We incorporate the proposed bit allocation and rate control algorithm into the MoMuSys MPEG-4 codec [28] and compare it to the VM8 rate control scheme [4]. The test QCIF video is "News" at 64 kbps with two objects as shown in Fig. 4. The frame rate is 10 fps. The buffer size is set to be 6400 bits, which is the target number of bits per frame. Fig. 6 shows the number of bits produced by each video frame when the proposed algorithm and the VM8 algorithm are applied. The actual coding bit rate is much closer to the target bit rate when the proposed algorithm is applied. The numbers of bits assigned to each video object are depicted in Fig. 7. Note that Object 2 (background) takes

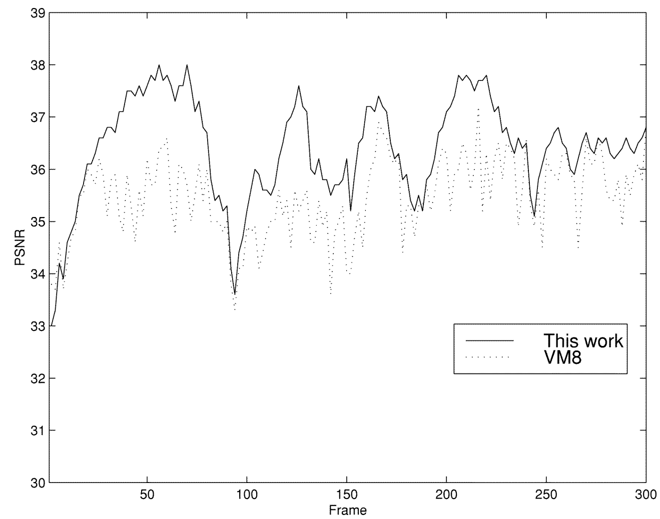


Fig. 8. PSNR of each frame when (solid line) the proposed algorithm and (dotted line) the VM8 algorithm are applied to the MPEG-4 codec.

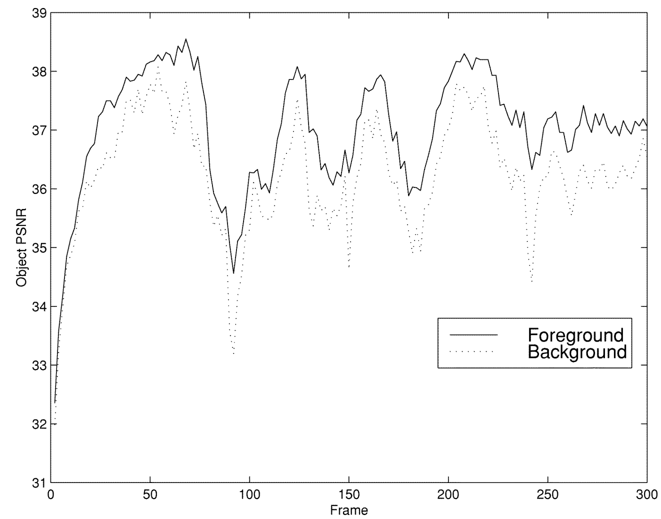


Fig. 9. PSNR of each VOP in the foreground and background video objects when their distortion weights are set to be 1.1 and 0.9 in the bit allocation scheme.

more bits than Object 1 (foreground). This is because the major activity in the scene is from the dancers (center background) in Object 2. Fig. 8 shows the overall PSNR of each video frame. With the proposed bit allocation scheme, we see that the picture quality is significantly improved. The improvement, about 1.15 dB on average, achieved by the proposed algorithm is due to its more accurate R–D models and more robust rate control.

It is straightforward to extend the proposed bit allocation scheme to take into account the subjective quality. For example, in general, the user is more interested in the moving foreground objects in the scene. In our bit allocation and rate control scheme, we can assign more bits to these objects and code them with higher fidelity. This can be realized by introducing a distortion weight w_i for each object into the objective function as follows:

$$\min_{\rho_i} \quad \sum_{i=1}^{K} w_i \sigma_i^2 e^{-\alpha_i(1-\rho_i)} \cdot N_i \tag{24}$$

$$\text{s.t.} \quad \sum_{i=1}^{K} \theta_i \cdot (1-\rho_i) \cdot N_i = R_T. \tag{25}$$

In this case, the optimum number of bits assigned to each object is given by

$$R_i = \xi_i w_i N_i \ln \frac{\sigma_i^2}{\xi_i} + \frac{\xi_i w_i N_i}{\sum_{i=1}^{K} \xi_i w_i N_i} \left(R_T - \sum_{i=1}^{K} \xi_i w_i N_i \ln \frac{\sigma_i^2}{\xi_i} \right). \tag{26}$$

Obviously, the objects of interest should have relatively larger weights to guarantee that they are coded with less distortion. In Fig. 9, we plot the PSNR of each VOP when the distortion weights of the foreground and background objects are set to be 1.1 and 0.9, respectively. We can see that with weighted bit allocation, the object of interest has better quality.

VII. R–D Optimized Coding for H.263

In the above section, the optimum bit allocation scheme developed in Section V is applied to MPEG-4 coding at the video object level. As we can see in this section, it can also be applied to video coding at the macroblock level. In a QCIF video frame, there are 99 16×16 macroblocks. We observe that it is not efficient to apply the bit allocation scheme directly to distribute the bit budge among these 99 macroblocks for the following two reasons. First, a very large number of data sources implies on average very few bits allocated to each source. Effective distribution of a smaller number of bits requires more accurate R–D models used in the bit allocation scheme. In others words, a very small R–D modeling error may cause significant performance degradation of the bit allocation. Second, these 99 macroblocks may have a wide range of R–D characteristics. For example, some macroblocks may be very active while others may be inactive. (Their coefficients are very close to zeros.) This will introduce strong singularity into the bit allocation scheme, which often results in negative bits assigned to those inactive macroblocks.

To improve the efficiency and robustness of the optimum bit allocation scheme, we first classify the 99 macroblocks into three classes according to their activity measures. The bit allocation scheme is then employed to distribute the bit budget among these three classes instead of among the 99 macroblocks. We observe that, after classification and grouping, the singularity introduced to the allocation process is significantly reduced. From our extensive simulation experience, it appears that "three" is a good choice for the class number with which the optimum bit allocation operates most effectively and robustly.

A. Macroblock Classification

The activity measure we choose for macroblock classification is the variance of the macroblock, denoted by $\{\sigma_j^2 | 1 \le j \le N\}$, where N is the total number of macroblocks in the current video frame. In our classification scheme, we first rearrange all of the macroblocks in decreasing order according to their variances.

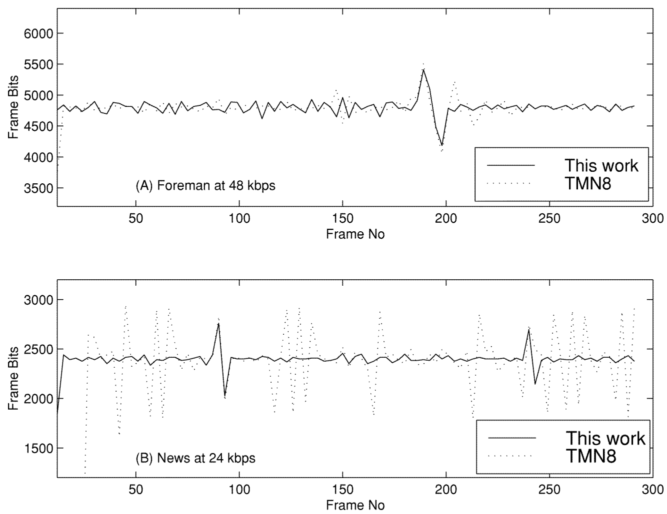


Fig. 10. Bits per frame when the proposed algorithm and the TMN8 algorithm are applied to the H.263 codec. "Foreman" at 48 kbps and "News" at 24 kbps.

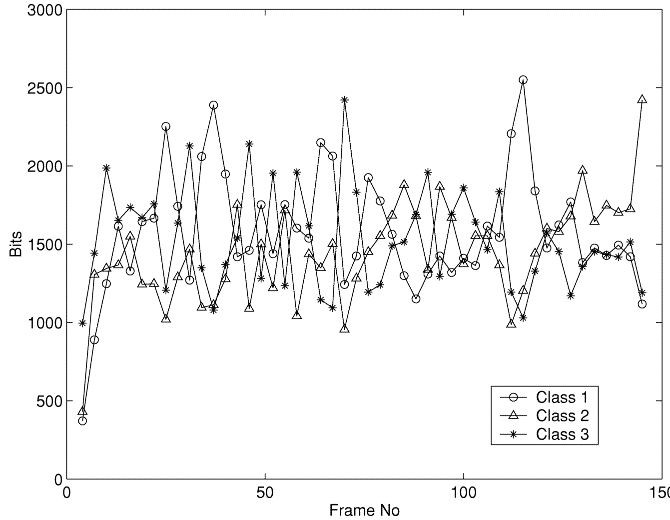


Fig. 11. Numbers of bits assigned to each macroblock class when the proposed bit allocation scheme is applied to H.263 coding of "Foreman."

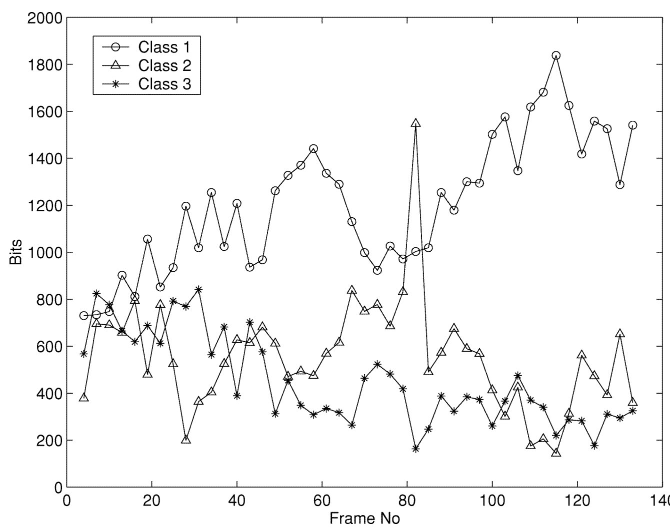


Fig. 12. Numbers of bits assigned to each macroblock class when the proposed bit allocation scheme is applied to H.263 coding of "News."

The first class S_1 consists of the top ten macroblocks which are the most active. The third class S_3 consists of the last 69 macroblocks which are most inactive. The remaining 20 macroblocks are placed into the second class S_2. The class sizes 10, 20, and 69 are chosen based on our experience. Certainly, more sophisticated classification scheme [26] can be applied to further improve the bit allocation and coding performance. However, for low computational complexity and implementation cost, we just use the above simple macroblock classification scheme in our bit allocation and rate control algorithm. After macroblock classification, each class is then treated as a separate input source. The parameters of the R–D models for each source can be determined by the same method as discussed in Section VI-A. To be more specific, we just treat each class of macroblocks as a generic video object as in MPEG-4 coding, and estimate the value of θ with (22) and (23). Once these two model parameters are obtained, with the optimum bit allocation scheme, the number of bits R_i assigned to each class is determined by (21). The ρ-RC rate control algorithm can be employed to achieve the bits target R_i for each class during the coding process.

B. Experimental Results

We incorporate the proposed bit allocation scheme and rate control algorithm into the H.263+ codec [18] and compare it with the TMN8 rate control scheme [5]. The two test QCIF videos are "Foreman" at 48 kbps and "News" at 24 kbps. The frame rate is fixed at 10 fps. Fig. 10 shows the number of bits produced by each video frame when the proposed algorithm and the TMN8 algorithm are applied. The numbers of bits assigned to each macroblock class are depicted in Figs. 11 and 12. The actual coding bit rate is shown to be much closer to the target bit rate when the proposed algorithm is applied, especially at lower bit rates. Fig. 13 shows the PSNR value of each video frame. With the proposed bit allocation scheme, the picture quality is significantly improved. Note that the TMN8 rate control algorithm has already included an optimum bit allocation scheme. But, the TMN8 bit allocation is based on the traditional q-domain R–D formulas. The improvement, about 0.8–1.0 dB, achieved by the proposed algorithm is due to our more accurate rate and distortion models.

The macroblock classification can also be combined with other functionalities, such as motion tracking [29] and region-of-interest coding [30], [31]. For example, we can classify the macroblocks in the current frame into the following classes: region of interest, background, and everything else in between. In this way, the macroblock classification is closely related to the user's interests and requirements. In the optimum bit allocation scheme, we can assign different distortion weights to different regions. In this way, the region of interest can be assigned more bits and coded with higher quality.

VIII. Concluding Remarks

In this paper, based on the ρ-domain R–D analysis framework, a distortion model is first developed in the ρ domain.

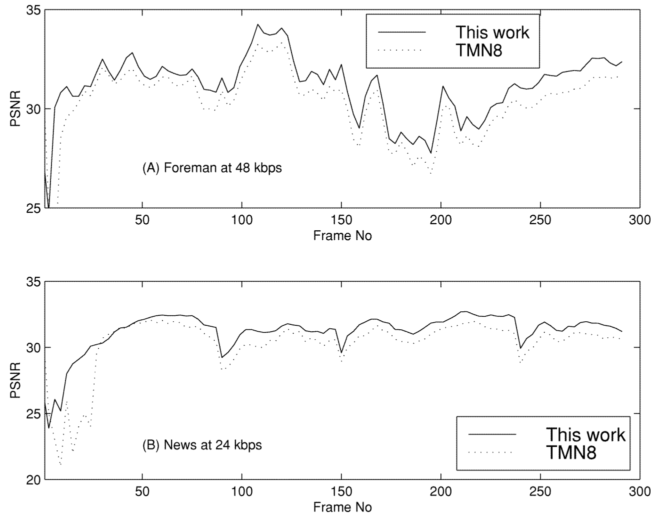


Fig. 13. PSNR of each frame when the proposed algorithm and the TMN8 algorithm are applied to the H.263 codec. "Foreman" at 48 kbps and "News" at 24 kbps.

Based on this distortion model and the linear rate model developed [1], an optimum bit allocation scheme is developed. It is applied to MPEG-4 video coding to distribute the bit budget among different video objects. It has also been applied to H.263 coding at the macroblock level when coupled with macroblock classification. The proposed bit allocation scheme extends the capability of the ρ-RC rate control algorithm. In addition, significant picture quality improvement is achieved due to the accurate R–D models developed in the ρ-domain. The proposed ρ-domain R–D analysis framework, bit allocation, and rate control scheme also have potential applications in other video coding scenarios, such as video transcoding, picture quality optimization in VBR coding for stored or streaming video under buffer constraint, R–D optimized adaptive frame rate selection, joint source-channel coding, etc.

REFERENCES

[1] 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," presented at the Int. Conf. Acoustics, Speech, and Signal Processing, Salt Lake City, UT, May 2001.

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

[3] 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.

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

[5] 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.

[6] 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.

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

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

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

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

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

[12] 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.

[13] K. H. Yang, A. Jacquin, and N. S. Jayant, "A normalized rate–distortion model for H.263-compatible codecs and its application to quantizer selection," in *Proc. 1997 Int. Conf. Image Processing*, Oct. 1997, pp. 41–44.

[14] B. Tao, H. A. Peterson, and B. W. Dickinson, "A rate–quantization model for MPEG Encoders," in *Proc. 1997 Int. Conf. Image Processing*, Oct. 1997, pp. 338–341.

[15] ITU-T, "Video coding for low bit rate communications,", ITU-T Recommendation H.263, version 1, Nov. 1995.

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

[17] Z. He, T.-H. Yu, and S. K. Mitra, "Blockwise zeros mapping image coding," in *Proc. Int. Conf. Image Processing*, Vancouver, BC, Canada, Sept. 2000.

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

[19] 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.

[20] J. J. Huang and P. M. Schultheiss, "Block quantization of correlated Gaussian random variables," *IEEE Trans. Commun. Syst.*, vol. CS-11, pp. 289–296, Sept. 1963.

[21] A. Segall, "Bit allocation and encoding for vector sources," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 162–169, Mar. 1976.

[22] B. Fox, "Discrete optimization via marginal analysis," *Manage. Sci.*, vol. 13, pp. 210–216, Nov. 1966.

[23] 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.

[24] R. L. Joshi, T. R. Fischer, and R. H. Bamberger, "Optimum classification in subband coding of images," in *Proc. Int. Conf. Image Processing*, vol. 2, Nov. 1994, pp. 883–887.

[25] H.-J. Lee, T. Chiang, and Y.-Q. Zhang, "Scalable rate control for MPEG-4 video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, pp. 878–894, Sept. 2000.

[26] R. L. Joshi, T. R. Fischer, and R. H. Bamberger, "Optimum classification in subband coding of images," in *Proc. Int. Conf. Image Processing*, vol. 2, Nov. 1994, pp. 883–887.

[27] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Norwell, MA: Kluwer, 1992, ch. 6.

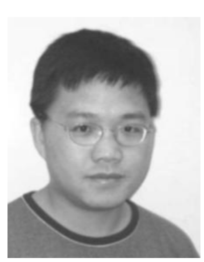
[28] MoMuSys codec, "MPEG4 verification model version 7," ISO/IEC JTC1/SC29/WG11 Coding of Moving Pictures and Associated Audio MPEG97, Bristol, U.K., Mar. 1997.

[29] D. Decarlo and D. Metaxas, "Optical flow constraints on deformable models with applications to face tracking," *Int. J. Comput. Vis.13p*, vol. 38, pp. 99–127, Aug. 2000.

[30] M. Kunt, A. Ikonomopoulos, and M. Kocher, "Second generation image coding techniques," *Proc. IEEE*, vol. 73, pp. 549–575, Apr. 1985.

[31] J. Benois-Pineau, D. Barba, N. Sarris, and M. G. Strintzis, "Video coding for wireless varying bit-rate communications based on area of interest and region representation," in *Proc. Int. Conf. Image Processing*, vol. 3, Santa Barbara, CA, Oct. 1997, pp. 555–558.

[32] ITU-T, "Video coding for low bit rate communications," ITU-T Recommendation H.263, version 2, Jan. 1998.



Zhihai He (M'01) received the B.S. degree from Beijing Normal University, Beijing, China, and the M.S. degree from Institute of Computational Mathematics, Chinese Academy of Sciences, Beijing, China, in 1994 and 1997 respectively, both in mathematics, and the Ph.D. degree from University of California, Santa Barbara, CA, in 2001, in electrical engineering.

In 2001, he joined Sarnoff Corporation, Princeton, NJ, as a Member of Technical Staff. His current research interests include image compression, video coding, network transmission, wireless communication, and embedded system design.

Dr. He received the 2001 IEEE Circuits and Systems Society CSVT Transactions Best Paper Award.

Sanjit K. Mitra (S'59–M'63–SM'69–F'74–LF'01) 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 in 1956 from Calcutta University, Calcutta, India, the M.S. and Ph.D. degrees in electrical engineering from the University of California, Berkeley, in 1960 and 1962, respectively, and an Honorary Doctorate of Technology degree from the Tampere University of Technology, Tampere, Finland, in 1987.

From June 1962 to June 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 at the University of California since then, first at the Davis campus and then at the Santa Barbara campus since 1977, as a Professor of Electrical and Computer Engineering, where he served as Chairman of the Department from July 1979 to June 1982. He has published over 550 papers in signal and image processing, 12 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 during 1996–1999. He is currently a member of the editorial boards of four journals. He is the recipient of the 1973 F.E. Terman Award and the 1985 AT&T Foundation Award of the American Society of Engineering Education, the 1989 Education Award, and the 2000 Mac Van Valkenburg Society Award of the IEEE CAS Society, the Distinguished Senior U.S. Scientist Award from the Alexander von Humboldt Foundation of Germany in 1989, the 1996 Technical Achievement Award, and the 2001 Society Award of the IEEE SP Society, the IEEE Millennium Medal in 2000, and the McGraw-Hill/Jacob Millman Award of the IEEE Education Society in 2001. He is the co-recipient of the 2000 Blumlein–Browne–Willans Premium of the Institution of Electrical Engineers (London), the 2001 IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY Best Paper Award, and the 2002 Technical Achievement Award of the European Association for Signal Processing (EURASIP). He is an Academician of the Academy of Finland and a Corresponding Member of the Croatian Academy of Sciences and Arts. He is a Fellow of the AAAS, and SPIE, and a member of EURASIP and ASEE.