Journal of Circuits, Systems, and Computers
Vol. 14, No. 6 (2005) 1157–1171

LOSS AWARE RATE ALLOCATIONS IN H.263 CODED VIDEO TRANSMISSIONS

XIAO SU

*Computer Engineering Department, San Jose State University,
One Washington Square, San Jose, CA 95192, USA*

BENJAMIN W. WAH

*Coordinated Science Laboratory,
Department of Electrical and Computer Engineering,
University of Illinois at Urbana-Champaign,
Urbana, IL 61801, USA*

Revised 6 August 2005

For packet video, information loss and bandwidth limitation are two factors that affect video playback quality. Traditional rate allocation approaches have focused on optimizing video quality under bandwidth constraint *alone*. However, in the best-effort Internet, packets carrying video data are susceptible to losses, which need to be reconstructed at the receiver side. In this paper, we propose loss aware rate allocations in both group-of-block (GOB) level and macroblock level, given that certain packets are lost during transmissions and reconstructed using simple interpolation methods at the receiver side. Experimental results show that our proposed algorithms can produce videos of higher quality when sent over lossy Internet.

Keywords: Error concealment; interpolation-based reconstruction; multi-description coding (MDC); real-time multimedia in the Internet; rate allocation.

1. Introduction

Although video coding and transmission have attracted much attention from research community, it remains to be a challenging topic. Two difficult issues involved are bandwidth constraints and information loss. In the literature, schemes can be found to address either of the two problems.

For information loss, there are sender-based,[1–6] receiver-based,[7–11] or sender–receiver-based schemes to recover from losses.[12–14] In particular, multiple-description coding (MDC) is an attractive approach for video streaming on the Internet because it greatly improves the error-resilience of coded bit streams. It divides video data into equally important streams such that the decoding quality using any subset is acceptable, and better quality is obtained by more descriptions. However, such schemes normally assume bandwidth is infinitely available.

For bandwidth constraints, there are rate allocation and adaptation schemes to optimize video quality within a limited rate budget.[15–22] Again, such techniques work best for error free environment.

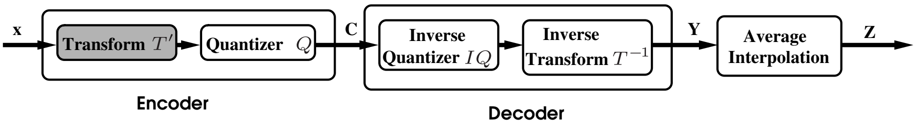
Different from existing schemes that deal with rate allocations under lossless conditions, we study rate allocations for lossy transmissions in which parts of a bit stream may get lost and need to be reconstructed. To our best knowledge, no efforts have been made to tackle the problem when sender employs certain robust coding algorithms, such as MDC. The proposed work fills in this gap. In such a setting, the design of rate allocation schemes is closely related to those of multiple-description coding at a sender and the reconstruction algorithm employed at a receiver.

To facilitate discussions, let us first list the notations to be used in the paper in Table 1. The general problem to be studied is as follows: *given the available bandwidth R, how do we design an* $\underline{\text{MDC}}$ *in order to minimize* $\underline{\text{reconstruction } \mathcal{E}_r}$*, subject to the rate constraint:* $r \leq R$*?* In the above statement, r is the actual rate (in bit per second) that the video signals are coded, and the reconstruction error $\mathcal{E}_r$ refers to the distortion between the original video signals before applying MDC and the recovered signals after decoding and reconstruction. One wide-adopted metric to measure $\mathcal{E}_r$ is Mean Squared Error (MSE) that calculates the average error between the original and the reconstructed pixel values.

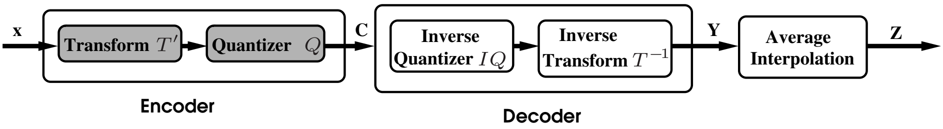
Figure 1 illustrates the basic building blocks of the encoding and decoding descriptions in MDC. Among the steps, Transform T and Quantizer Q are two very important components that can greatly affect video playback quality. However, in lossy situations, the original Transform T and Quantizer Q are not designed

Table 1. Notations to be used in the paper.

Notation	Definition
$\mathcal{E}_r$	Reconstruction error, i.e., difference between the original and the reconstructed videos
R	Rate budget, i.e., available video coding rate (in bit per second)
r	Actual video coding rate (in bit per second)
Q_i	Quantization factor for the ith frame
q_i	Quantization factor for the ith GOB in a frame
s_i	Quantization factor for the ith coefficient in a block
$D(\cdot)$	Rate-distortion function. $D_i(x_i)$ represents distortion when video coding rate is x_i
X	Vector of pixels in an original block
X'	Vector of pixels in a reconstructed block
Y	Vector of original transformed coefficients in a block
Y'	Vector of reconstructed transformed coefficients in a block (after quantization)
d_i	Quantization error for the ith coefficient in a block
σ_i^2	Variance of the ith coefficient in a block
c_i	The ith coefficient in a block
R_o	The bit rate resulted from the original quantization method
R_s	The bit rate resulted from the scaled quantization method
$PSNR_o$	Peak-Signal-to-Noise-Ratio of the original quantization method
$PSNR_s$	Peak-Signal-to-Noise-Ratio of the scaled quantization method



(a) Optimized Reconstruction-Based DCT (ORB-DCT).



(b) H.263 codec with modified transform T and Quantizer Q.

Fig. 1. Modified H.263 codec for reconstruction purpose.

for optimal *reconstruction* performance. In Ref. 23, we have proposed Optimized Reconstruction-Based DCT (ORB-DCT) that modified only Transform T, but not Quantizer Q, as illustrated in Fig. 1(a). To add rate constraints, we need to modify both T and Q, as illustrated in Fig. 1(b). However, such a formulation involving quantization module Q is a constrained integer optimization problem and is not solvable in a closed-form. Therefore, in this paper, we discuss heuristic approaches to address this problem.

Modifying Quantizer Q results in different rate allocations in the frame, group-of-block (GOB), and macroblock levels. Figure 2 shows how rate control and allocations can be done in each layer. At the top frame level, rate allocations can be achieved by assigning distinct Q_i's to frames. At the GOB level, rate allocations can be done by assigning different q_i's to blocks within the GOB. The assignment of q_i's overrides the default quantization choice set at the frame level. At the macroblock level, rate allocations can be done by applying different s_i's to coefficients within a macroblock. Again, the value of s_i overrides the quantization choice set at the

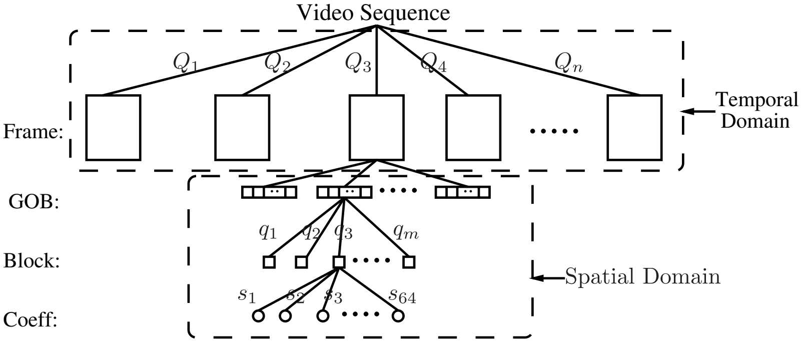


Fig. 2. Rate allocation and control problems in H.263.

GOB level. In this paper, we focused on two spatial domain schemes implemented in GOB- and macroblock-levels.

The paper is organized as follows. In Sec. 2, we discuss reconstruction-based rate allocation among macroblocks in GOB-level. In Sec. 3, we proposed loss aware quantization schemes for individual coefficients within a macroblock. Section 4 concludes the paper.

2. Reconstruction-Based Rate Allocation Among Blocks in a GOB

As a GOB consists of a sequence of macroblocks, and if the total rate allocated to this GOB is constrained by a budget R, the question is how to choose quantization factors among macroblocks within the GOB in order to maximize reconstruction performance, subject to the rate constraint. To facilitate future discussions, we define notations to be used in Table 1.

Let us start by reviewing the solution to this problem, without considering the fact that video signals may get lost and need to be reconstructed. The classical solution to this problem is based on the following theorem.

Theorem 1.[18] Given that the rate-distortion functions of macroblocks, $D_i(x_i), i = 1, 2, \ldots, n$, are convex, the rate allocation vector $(r_1, r_2, \ldots, r_n)$ is the solution to*:*

$$\min \sum_i D_i(x_i)$$
$$s.t. \sum_i x_i \leq R$$

if and only if the following condition satisfies:

$$\left(\frac{\partial D_1}{\partial x_1}\right)_{r_1} = \left(\frac{\partial D_2}{\partial x_2}\right)_{r_2} = \cdots = \left(\frac{\partial D_n}{\partial x_n}\right)_{r_n}.$$

The proof can be found in Ref. 18, and the discrete version of the theorem can be found in Ref. 21. Essentially, the derivatives $(\partial D_i/\partial x_i)_{r_i}$, $i = 1, 2, \ldots, n$, are the slopes of lines tangent to the rate-distortion (R-D) curves of the macroblocks coded at rates r_i, $i = 1, 2, \ldots, n$. For this reason, the algorithm implementing the theorem is normally referred to as "constant slope optimization". The intuitive idea behind the algorithm is very simple. At those points with constant slope, all the macroblocks operate at the same marginal return for an extra bit in the rate-distortion trade-off. In other words, if we reduce one bit for macroblock i, and spend it on another macroblock j (to maintain the same bit rate), then the reduction in distortion of macroblock j would be equal to the increase in distortion of macroblock i. For this reason, there is no allocation that is more efficient for this rate budget.

This theorem establishes the necessary and sufficient conditions for optimal rate allocations among macroblocks. To apply the theorem, one needs to verify an important assumption, i.e., the R-D curve for each individual block is convex. It has been found that conventional single description coders (SDC) generate convex R-D curves, but no results have been reported about MDC coders with reconstructions. Next, we empirically establish the properties of the R-D curves for MDC coders with reconstruction. Please note that in MDC setting, the distortion is calculated between the decompressed *and reconstructed* signals and the original signals.

To this end, we first modified the MDC-based H.263 codec in such a way that the reconstruction quality after interpolation and the corresponding bit rate spent on each macroblock were saved for each description, for a given quantization choice. Then, we iterated through all possible quantization choices, i.e., $2, 3, \ldots, 31$, and obtained 30 rate-distortion pairs, that resulted in a rate-distortion (R-D) curve for each macroblock.

From the experiments, we have found that all the intra-coded macroblocks and a majority of the inter-coded macroblocks have convex R-D curves. Some inter-coded macroblocks have nonconvex R-D curves due to their complex dependencies on the R-D curves of their reference macroblocks. To save space, we only show the R-D curves of four randomly chosen intra-coded macroblocks and four inter-coded macroblocks from three test video sequences: *missa*, *football*, and *akiyo*. In Fig. 3, the first row shows the R-D curves of intra-blocks from *missa*, and the second row shows the curves of inter-blocks from *missa*, followed by those of intra- and inter-blocks from *football* and *akiyo*. In these plots, rate is measured in bytes, and distortion is calculated in terms of mean squared error. Although for some video sequences, their curves are not convex in certain small local regions, convexity is still observed in most parts of all the R-D curves. We can observe the same trend in the R-D curves of other test sequences that are not included here due to space constraints. As a result, we conclude that the R-D relationship for reconstructed macroblocks in MDC is approximately convex; therefore, previous approaches that address optimal allocations among macroblocks can still be applied in MDC with reconstruction.[18,21,24]

3. Design of Quantization Matrices for MDC

H.263 uniformly quantizes every coefficient in a block by applying the same quantization factor q. Intuitively, this simple scheme is not optimal because it does not exploit the characteristics of individual coefficients. The objective of our work is to improve its performance for MDC by assigning proper quantization factors to different coefficients.

As quantization is done in the coefficient domain after DCT transform, we need to first relate errors introduced in the coefficient domain to those observed in the pixel domain. Let X and X' denote the original and the reconstructed blocks of pixels, and Y and Y' be the corresponding original and reconstructed blocks of transformed coefficients, we have the following relationship between the errors in

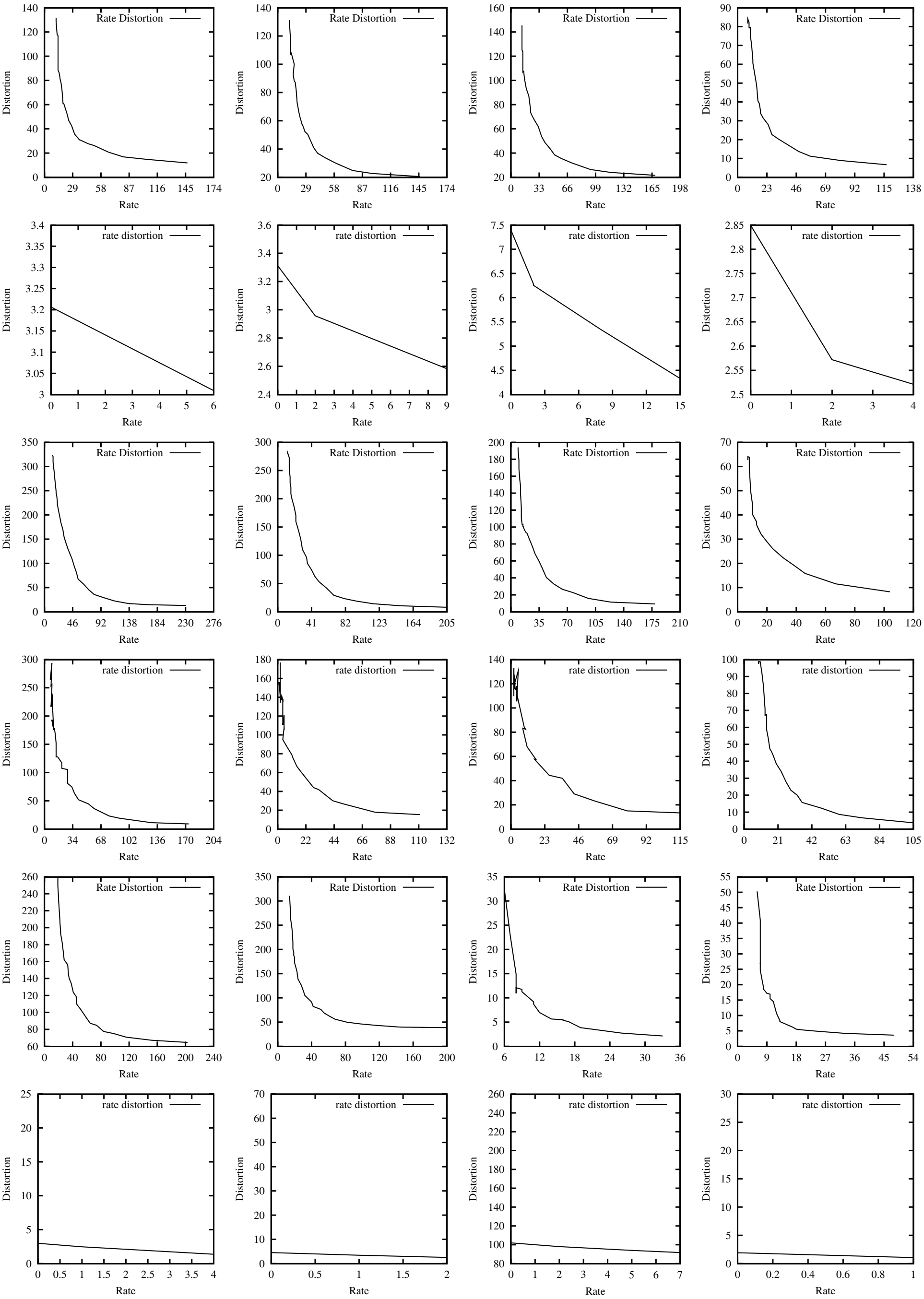


Fig. 3. Rate-distortion curves of four randomly chosen macroblocks from a I-frame of *missa*, P-frame of *missa*, I-frame of *football*, P-frame of *football*, I-frame of *akiyo*, and P-frame of *akiyo*.

these two domains:

$$\begin{aligned}
\|Y - Y'\|^2 &= \|TX - TX'\|^2 && (1)\\
&= (X - X')^T T^T T (X - X') && (2)\\
&= (X - X')^T (X - X') \quad \text{if and only if } T^T T = I && (3)\\
&= \|X - X'\|^2 \,. && (4)
\end{aligned}$$

Equations (2) and (3) hold if and only if T is an orthonormal matrix. It is easy to verify that DCT is an orthonormal transform; therefore, the energy of quantization errors in DCT transform coefficients is equal to that of image pixels. This is a useful property because it implies that our efforts to reduce quantization errors are equally reflected in the pixel domain as well.

To find the quantization factor for each coefficient in a block, we need to first find the number of bits to be allocated to each coefficient and then map this rate budget to quantization factor. For the first step, we can formulate the block-level rate allocation as a constrained optimization problem as follows[25]:

$$\min \mathcal{E}_r = \frac{1}{n}\sum_{i=1}^{n} d_i^2 \,, \tag{5}$$

$$s.t. \sum_{i=1}^{n} R_i = R \,. \tag{6}$$

In the above formulation, we assume that ith coefficient is quantized to R_i bits, and the resulting quantization error is d_i. There are n coefficients in a block. In H.263, n is equal to 64 since each block is 8 by 8. The optimal solution can be obtained in terms of the rate budget, R, the number of coefficients within the macroblock, n, and coefficient variances[25]: $\sigma_i, i = 1, 2, \ldots, n$:

$$R_i = \frac{R}{n} + \frac{1}{2}\log_2 \frac{\sigma_i^2}{\sqrt[n]{\sigma_1^2 \sigma_2^2 \cdots \sigma_n^2}} \,, \quad i = 1, 2, \ldots, n \,. \tag{7}$$

The equation indicates that bit allocation should be done based on coefficient variances. If all the coefficients have equal variances, i.e., $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2$, then, the best way is to assign R/n bits to each coefficient. On the other hand, if the variance of a certain coefficient i, σ_i^2, is greater (or smaller) than the geometrical average of the variances, then, the number of bits allocated to this coefficient should be greater (or smaller) than R/n, the average number of bits for each coefficient.

Although mathematically elegant, it is difficult to directly apply this closed-form solution in practice. First, real-time estimations of the variance for each coefficient results in an increase in both memory requirement and computational time. We need to set aside a frame buffer to facilitate variance estimation, and this doubles the memory requirement in encoding. To understand its computational overhead, let us do a rough analysis on the number of computations needed to estimate coefficient variance. Suppose there are m blocks in each frame, then, we need to estimate the variances for 64 coefficient bands, $c_i, i = 1, 2, \ldots, 64$ among the m blocks as follows:

$$\bar{c} = \frac{1}{m} \sum_{i=1}^{m} c_i \,, \tag{8}$$

$$\sigma_i^2 = \frac{1}{m} \sum_{i=1}^{m} \|c_i - \bar{c}\|^2 \,. \tag{9}$$

To calculate the average coefficient value, we need to perform m additions and one division. To calculate the coefficient variance for a single coefficient, σ_i^2, we need m additions, m subtractions, m multiplications and one division. Therefore, we need $4m + 2$ computations to estimate the variance for a single coefficient, and $64(4m + 2)$ to calculate all the variances. For a small CIF video sequence where the frame dimension is 352×288, m is equal to $1584 = (352/8 \times 288/8)$. This results in a total number of 405 632 computations to be performed to estimate coefficient variance! Compared to the original quantization method that needs just one division to quantize each coefficient and 101 376 computations to quantize all the coefficients, this results in a four-fold increase in computational time.

Second, due to the nonstationarity of video frames, the variance of a coefficient changes from frame to frame; hence, a quantization matrix needs to be sent for each frame, leading to additional bit overhead that may not justify the bit savings resulted from this approach.

Third, the *largest obstacle* to the application of this formula is that the relationship between rate and quantization factors cannot be derived in advance due to the zigzag ordering and variable length coding employed. As both of them have large impact on the resulted bit rate but cannot be formulated in closed-form, it is difficult to find the optimal quantization choices given the knowledge of this optimal bit allocation vector.

Equation (7), however, still provides guidelines for designing macroblock-level quantization schemes. Basically, it suggests that coefficients should be quantized according to their variances in the way that if the variance of a certain coefficient, σ_i^2, is greater (or smaller) than the geometrical average of the variances, then the number of bits allocated to this coefficient should be greater (or smaller) than R/n, the average number of bits for each coefficient. To develop practical MDC-based quantization schemes, our first step is to study how the MDC process changes the variances of individual coefficients. For this purpose, we group coefficients from a video frame into 64 bands by putting coefficients with the same coordinate (i, j) in a transformed block, $i, j = 1, 2, \ldots, 8$, into the same band, and calculate the variances of coefficients within each band. We do this separately for intra-coded and inter-coded frames because they have different inputs: intra-coded frames code the original pixel values, whereas inter-coded frames code the residual signals computed from the current and its reference frames.

Table 2 shows the ratio of coefficient variances of a horizontally-interleaved MDC system as compared to those of a SDC system, for three CIF format sequences

Table 2. Ratio of coefficient variances of a horizontally-interleaved MDC system compared to those of a SDC system, for intra-coded and inter-coded blocks from *missa*, *football*, *boxing*, *akiyo*, and *coastguard*, respectively.

	Intra-coded block								Inter-coded block							
missa	$\boxed{0.95}$	2.01	3.35	4.14	4.09	2.85	2.56	1.61	1.88	1.51	1.44	2.00	2.59	1.54	2.60	1.39
	$\boxed{0.84}$	1.15	2.43	3.39	2.62	2.62	2.59	1.38	1.25	1.30	1.46	1.65	2.31	2.07	1.51	1.21
	$\boxed{0.68}$	1.50	1.16	1.58	2.08	1.99	2.87	1.36	1.20	1.14	1.37	1.80	2.16	1.63	1.87	1.29
	$\boxed{0.60}$	1.27	1.11	1.91	1.34	1.75	1.89	1.22	$\boxed{0.87}$	1.03	1.20	2.05	1.19	1.73	1.66	1.41
	$\boxed{0.58}$	1.12	1.18	1.22	1.34	1.13	2.32	1.05	$\boxed{0.73}$	1.50	1.45	1.38	1.03	1.39	1.79	1.06
	$\boxed{0.51}$	1.46	1.26	1.34	1.24	1.22	1.73	1.02	$\boxed{0.71}$	1.05	1.10	1.29	$\boxed{0.97}$	$\boxed{0.93}$	1.62	1.30
	$\boxed{0.06}$	$\boxed{0.87}$	1.18	$\boxed{0.78}$	1.27	$\boxed{0.33}$	$\boxed{0.80}$	1.16	$\boxed{0.05}$	$\boxed{0.91}$	$\boxed{0.88}$	1.04	1.46	0.41	$\boxed{0.89}$	5.23
	$\boxed{0.08}$	$\boxed{0.86}$	1.48	1.03	1.14	$\boxed{0.46}$	1.07	1.23	$\boxed{0.06}$	$\boxed{0.90}$	1.35	1.33	1.30	$\boxed{0.49}$	1.01	3.27
football	$\boxed{0.92}$	2.50	2.64	1.91	2.16	2.62	5.37	9.41	1.50	1.69	2.12	2.25	1.88	4.06	5.56	9.51
	$\boxed{0.79}$	1.21	2.07	1.77	2.85	3.61	5.19	8.26	1.00	1.13	1.32	1.68	2.34	4.36	8.81	13.85
	$\boxed{0.66}$	1.02	1.45	2.25	2.93	3.94	4.63	7.67	$\boxed{0.83}$	1.02	1.06	1.64	2.65	4.20	9.01	14.18
	$\boxed{0.59}$	$\boxed{0.81}$	1.45	1.98	3.51	4.42	5.16	6.56	$\boxed{0.73}$	$\boxed{0.91}$	1.14	1.64	2.17	3.73	9.43	14.76
	$\boxed{0.66}$	$\boxed{0.91}$	1.76	1.91	3.10	5.09	6.15	7.47	$\boxed{0.85}$	$\boxed{0.88}$	1.35	1.36	2.85	6.12	8.68	13.24
	$\boxed{0.66}$	$\boxed{0.99}$	2.01	2.52	3.83	5.27	6.94	7.87	$\boxed{0.67}$	1.24	1.64	1.83	2.94	5.86	8.64	14.54
	$\boxed{0.65}$	1.11	1.49	2.48	4.15	6.36	7.11	12.12	$\boxed{0.62}$	1.38	1.35	1.22	3.62	9.52	10.97	10.20
	$\boxed{0.59}$	$\boxed{0.96}$	1.41	1.62	3.47	8.09	11.64	13.29	$\boxed{0.69}$	$\boxed{0.87}$	1.12	1.16	2.22	5.35	10.18	17.76
boxing	$\boxed{0.93}$	1.71	1.81	2.38	4.14	4.29	6.87	12.06	1.03	2.40	2.46	2.00	1.95	4.84	5.80	14.38
	$\boxed{0.86}$	1.33	1.41	1.48	2.19	2.93	12.53	26.26	$\boxed{0.80}$	1.39	1.67	1.54	2.03	4.70	7.55	14.99
	$\boxed{0.85}$	1.41	1.14	1.23	2.11	3.43	9.65	21.26	$\boxed{0.76}$	$\boxed{0.57}$	1.04	1.76	1.79	3.70	6.79	19.15
	$\boxed{0.75}$	1.26	1.21	1.29	1.81	4.32	7.51	20.18	1.13	$\boxed{0.41}$	$\boxed{0.88}$	1.06	1.48	3.13	5.24	23.38
	$\boxed{0.77}$	1.37	$\boxed{0.88}$	$\boxed{0.81}$	2.16	3.51	10.31	22.09	1.05	$\boxed{0.98}$	1.28	$\boxed{0.86}$	1.49	2.79	6.34	15.59
	$\boxed{0.74}$	1.09	1.06	1.29	1.73	3.28	6.38	12.34	$\boxed{0.71}$	$\boxed{0.93}$	1.02	1.14	1.37	2.36	5.40	11.11
	$\boxed{0.67}$	$\boxed{0.93}$	$\boxed{0.80}$	1.58	1.69	4.07	3.00	7.12	$\boxed{0.79}$	$\boxed{0.85}$	1.04	1.18	1.20	2.51	4.21	9.12
	$\boxed{0.71}$	$\boxed{0.88}$	$\boxed{0.97}$	$\boxed{0.76}$	$\boxed{0.52}$	2.97	3.09	6.60	$\boxed{0.73}$	$\boxed{0.16}$	$\boxed{0.22}$	$\boxed{0.36}$	$\boxed{0.55}$	2.12	4.59	7.69
akiyo	$\boxed{0.94}$	1.62	2.15	1.65	2.28	3.65	6.57	16.70	$\boxed{0.95}$	1.06	$\boxed{0.85}$	1.18	1.72	1.73	2.01	3.97
	$\boxed{0.76}$	1.38	$\boxed{0.81}$	$\boxed{0.65}$	2.06	2.45	4.91	20.84	1.14	1.08	1.13	1.37	1.69	1.70	1.81	2.66
	$\boxed{0.69}$	1.12	$\boxed{0.98}$	$\boxed{0.75}$	1.03	1.87	4.45	16.49	$\boxed{0.99}$	1.04	1.06	1.16	1.19	1.44	1.58	3.53
	$\boxed{0.62}$	$\boxed{0.74}$	1.30	1.28	1.88	1.30	2.39	6.22	1.12	1.41	$\boxed{0.96}$	$\boxed{0.79}$	1.23	1.45	1.50	2.93
	$\boxed{0.38}$	$\boxed{0.82}$	$\boxed{0.95}$	1.92	1.25	2.00	2.97	4.93	1.10	1.13	1.05	$\boxed{0.82}$	1.10	1.48	1.85	2.91
	$\boxed{0.55}$	$\boxed{0.31}$	$\boxed{0.74}$	1.25	3.12	2.69	3.54	7.67	1.03	$\boxed{0.88}$	$\boxed{0.76}$	$\boxed{0.91}$	2.10	1.78	2.54	4.73
	$\boxed{0.56}$	$\boxed{0.40}$	$\boxed{0.30}$	1.09	2.58	6.88	2.87	5.98	$\boxed{0.96}$	$\boxed{0.84}$	1.23	$\boxed{0.80}$	1.57	2.84	2.61	6.63
	$\boxed{0.43}$	$\boxed{0.55}$	$\boxed{0.40}$	1.21	$\boxed{0.91}$	4.64	$\boxed{0.93}$	19.92	$\boxed{0.88}$	$\boxed{0.55}$	$\boxed{0.49}$	1.08	1.63	5.81	3.38	16.56
coastguard	$\boxed{0.92}$	2.43	1.24	2.67	5.63	1.31	15.62	20.19	$\boxed{0.87}$	1.51	1.17	1.40	1.85	3.23	2.37	7.20
	$\boxed{0.90}$	1.35	1.36	1.60	1.60	4.19	5.64	14.89	$\boxed{0.88}$	1.07	1.12	1.42	1.88	2.72	6.18	18.12
	$\boxed{0.79}$	1.37	1.71	1.52	2.52	2.94	5.89	13.51	$\boxed{0.91}$	1.04	1.28	1.38	1.99	1.89	4.36	9.09
	$\boxed{0.86}$	1.07	$\boxed{0.97}$	1.04	1.64	3.07	9.42	21.87	$\boxed{0.96}$	1.19	$\boxed{0.97}$	1.38	1.85	2.35	6.27	13.06
	$\boxed{0.81}$	1.03	$\boxed{0.92}$	2.06	1.39	2.99	4.22	12.83	$\boxed{0.91}$	$\boxed{0.88}$	1.25	1.42	1.50	2.03	4.24	19.45
	$\boxed{0.70}$	1.10	1.29	1.14	1.16	2.76	6.43	23.45	$\boxed{0.83}$	1.12	1.30	$\boxed{0.93}$	1.79	2.48	5.66	19.75
	$\boxed{0.73}$	$\boxed{0.86}$	1.18	1.29	2.12	2.31	6.22	13.45	$\boxed{0.81}$	$\boxed{0.94}$	1.30	1.10	2.09	2.37	4.99	17.76
	$\boxed{0.70}$	1.05	$\boxed{0.95}$	1.57	1.70	2.66	3.60	19.49	$\boxed{0.81}$	1.26	1.08	1.67	1.59	2.45	5.55	15.39

(*missa*, *football*, and *boxing*) and two QCIF format sequences (*akiyo* and *coastguard*). The coefficients having smaller variances after MDC are circled in ovalboxes.

The results tell us that the variances in the upper right part of a coefficient block tend to increase after MDC, and those in the lower left tend to decrease after MDC.

This is not surprising because horizontal (resp. vertical) frequency components are likely to increase (resp. decrease) after horizontal partitioning, and the coefficients in the upper right (resp. lower left) triangle are the ones that capture horizontal (resp. vertical) frequencies. As we know that coefficients with large variances need to be quantized more finely than those with smaller variances, our observation motivates the following quantization scheme for MDC:

$$Q_{i,j} = \begin{cases} \alpha Q & i \geq j \\ \beta Q & i < j \end{cases} \quad \alpha \geq 1\,, \ \ \beta \leq 1\,,$$

where $Q_{i,j}$ is the quantization factor to be used for the coefficient of row i and column j, α and β are scaling parameters, and Q is the original quantization choice for this block. To choose suitable α and β, we have evaluated the following combination of choices: $\alpha = 1.0, 1.05, \ldots, 1.2$ and $\beta = 0.7, 0.75, \ldots, 1$, for each video sequence.

The best results along with the parameters and the comparisons with the original quantization scheme can be found in Table 3. Here, R_s (resp. R_o) represents the bit rate resulted from the scaled (resp. original) quantization, measured in bytes, and $PSNR_s$ (resp. $PSNR_o$) denotes PSNR values for the scaled (resp. original) approach. From the results on $\Delta PSNR$ $(= PSNR_s - PSNR_o)$ and $\Delta R/R$ $(= (R_s - R_o)/R_o)$, we can see that the modified quantization scheme lead to better PSNRs and 1–10% savings in bit rates for *missa*, *football*, *boxing*, *akiyo*, and *river*, and comparable R-D results for *coastguard*.

In our approach, since the same scaling factors are used throughout a video sequence, there is no overhead in bit rate when compared to approaches that need to send frame-based quantization matrices to decoders. Furthermore, the estimations of variances and scaling factors, α and β, do not add much extra complexity in real-time encoding because they can be done offline.

A natural question arises as to how our proposed quantization algorithm increases computational time in real-time encoding and decoding, since the quantization and de-quantization processes become floating point operations after scaling. To this end, we computed encoding time with the original quantization ($enct_o$), encoding time with the proposed quantization ($enct_s$), decoding time with the original quantization ($dect_o$), and decoding time with the proposed quantization ($dect_s$) and reported them in Table 4. These numbers were calculated as the averages of 100 experimental runs. In each run, we recorded time to encode and decode 90 frames of each sequence, respectively. The experiments were done on a Pentium-III PC with 1.8 GHz CPU and 512 MB memory.

In Table 4, we can see that the increase in computational time due to floating number operations in quantization and dequantization is negligible, less than 2%. This can be partly explained by the fact that both quantization and dequantization only take a very small fraction of time in the encoding and decoding processes. In the literature, people have reported time profiling results of MPEG-2 and H.263 coding[26,27]: in encoding, around 85% time is spent on motion estimation and

Table 3. Comparisons of bit rates and PSNRs of scaled quantization and original quantization for *missa*, *football*, *boxing*, *akiyo*, *coastguard* and *river*, respectively.

Quant factor	R_o	$PSNR_o$	R_s	$PSNR_s$	$\Delta R/R_o$	$\Delta PSNR$
(a) *missa*: one description received ($\alpha = 0.9$, $\beta = 1.0$)						
4	520142	39.30	505814	39.37	$\boxed{-2.75\%}$	$\boxed{0.07}$
8	155992	37.87	140943	37.93	$\boxed{-9.65\%}$	$\boxed{0.06}$
12	81494	36.74	79162	36.89	$\boxed{-4.01\%}$	$\boxed{0.15}$
16	52638	35.96	51285	36.01	$\boxed{-1.27\%}$	$\boxed{0.05}$
20	38244	35.22	37814	35.26	$\boxed{-1.09\%}$	$\boxed{0.04}$
(b) *missa*: two descriptions received ($\alpha = 0.9$, $\beta = 1.0$)						
4	1041431	39.70	1007594	39.83	$\boxed{-3.25\%}$	$\boxed{0.13}$
8	312739	37.94	283405	38.04	$\boxed{-9.38\%}$	$\boxed{0.10}$
12	161445	36.79	158298	36.93	$\boxed{-1.95\%}$	$\boxed{0.14}$
16	104896	35.94	102589	36.00	$\boxed{-2.20\%}$	$\boxed{0.06}$
20	75707	35.21	75190	35.22	$\boxed{-0.68\%}$	$\boxed{0.01}$
(c) *football*: one description received ($\alpha = 0.95$, $\beta = 1.05$)						
4	1370410	34.02	1313080	34.08	$\boxed{-4.18\%}$	$\boxed{0.06}$
8	686309	31.80	664489	31.83	$\boxed{-3.18\%}$	$\boxed{0.03}$
12	429069	30.14	417638	30.15	$\boxed{-2.66\%}$	$\boxed{0.01}$
16	297435	28.94	292659	28.96	$\boxed{-1.61\%}$	$\boxed{0.02}$
20	222290	28.02	219458	28.05	$\boxed{-1.27\%}$	$\boxed{0.03}$
(d) *football:* two descriptions received ($\alpha = 0.95$, $\beta = 1.05$)						
4	2739181	35.60	2624672	35.73	$\boxed{-4.18\%}$	$\boxed{0.13}$
8	1371836	32.25	1328216	32.32	$\boxed{-3.18\%}$	$\boxed{0.07}$
12	857213	30.23	834977	30.28	$\boxed{-2.59\%}$	$\boxed{0.05}$
16	594469	28.88	585938	28.93	$\boxed{-1.44\%}$	$\boxed{0.05}$
20	443592	27.91	438298	27.96	$\boxed{-1.19\%}$	$\boxed{0.05}$
(e) *boxing*: one description received ($\alpha = 0.95$, $\beta = 1.05$)						
4	6433505	32.96	6202908	32.99	$\boxed{-3.58\%}$	$\boxed{0.03}$
8	3372707	31.37	3299831	31.40	$\boxed{-2.16\%}$	$\boxed{0.03}$
12	2248014	29.99	2215691	30.05	$\boxed{-1.44\%}$	$\boxed{0.06}$
16	1660040	28.86	1653551	28.90	$\boxed{-0.39\%}$	$\boxed{0.04}$
20	1302048	27.93	1301098	27.96	$\boxed{-0.07\%}$	$\boxed{0.03}$
(f) *boxing*: two description received ($\alpha = 0.95$, $\beta = 1.05$)						
4	12873887	35.16	12408811	35.32	$\boxed{-3.61\%}$	$\boxed{0.16}$
8	6744048	32.30	6599903	32.42	$\boxed{-2.14\%}$	$\boxed{0.12}$
12	4493241	30.37	4428943	30.47	$\boxed{-1.43\%}$	$\boxed{0.10}$
16	3316086	28.98	3303632	29.07	$\boxed{-0.38\%}$	$\boxed{0.09}$
20	2600602	27.90	2598681	27.97	$\boxed{-0.07\%}$	$\boxed{0.07}$

Table 3. (*Continued*)

Quant factor	R_o	$PSNR_o$	R_s	$PSNR_s$	$\Delta R/R_o$	$\Delta PSNR$
(g) *akiyo*: one description received ($\alpha = 0.9$, $\beta = 1.0$)						
4	166766	33.18	148173	33.24	$\boxed{-11.2\%}$	$\boxed{0.06}$
8	84907	32.08	81653	32.23	$\boxed{-3.83\%}$	$\boxed{0.15}$
12	49612	31.09	47693	31.26	$\boxed{-3.87\%}$	$\boxed{0.17}$
16	39536	30.25	35720	30.46	$\boxed{-9.65\%}$	$\boxed{0.21}$
20	29138	29.48	27075	29.58	$\boxed{-7.08\%}$	$\boxed{0.10}$
(h) *akiyo*: two descriptions received ($\alpha = 0.9$, $\beta = 1.0$)						
4	338150	36.09	297852	36.26	$\boxed{-11.9\%}$	$\boxed{0.17}$
8	166909	33.78	164726	34.04	$\boxed{-1.31\%}$	$\boxed{0.26}$
12	99810	32.10	94733	32.38	$\boxed{-5.09\%}$	$\boxed{0.28}$
16	78012	30.87	70867	31.20	$\boxed{-9.16\%}$	$\boxed{0.33}$
20	58506	29.87	52890	30.05	$\boxed{-9.60\%}$	$\boxed{0.18}$
(i) *coastguard*: one description received ($\alpha = 0.95$, $\beta = 1.05$)						
4	870368	32.74	838909	32.71	$\boxed{-3.61\%}$	−0.03
8	434397	30.85	421569	30.89	$\boxed{-2.95\%}$	$\boxed{0.04}$
12	268977	29.34	269558	29.40	0.02%	$\boxed{0.06}$
16	182715	28.22	180423	28.17	$\boxed{-1.25\%}$	−0.05
20	133379	27.34	131247	27.31	$\boxed{-1.60\%}$	−0.03
(j) *coastguard*: two descriptions received ($\alpha = 0.95$, $\beta = 1.05$)						
4	1732269	34.70	1670505	34.71	$\boxed{-3.57\%}$	$\boxed{0.01}$
8	864098	31.62	837878	31.62	$\boxed{-3.03\%}$	0.00
12	534677	29.65	523749	29.63	$\boxed{-2.04\%}$	−0.02
16	363121	28.35	358719	28.31	$\boxed{-1.21\%}$	−0.04
20	264748	27.37	261419	27.35	$\boxed{-1.26\%}$	−0.02
(k) *river*: one description received ($\alpha = 0.95$, $\beta = 1.0$)						
4	1370410	34.02	1313080	34.08	$\boxed{-4.18\%}$	$\boxed{0.06}$
8	686309	31.80	664489	31.83	$\boxed{-3.18\%}$	$\boxed{0.03}$
12	429069	30.14	417638	30.15	$\boxed{-2.66\%}$	$\boxed{0.01}$
16	297435	28.94	292659	28.96	$\boxed{-1.61\%}$	$\boxed{0.02}$
20	222290	28.02	219458	28.05	$\boxed{-1.27\%}$	$\boxed{0.03}$
(l) *river*: two descriptions received ($\alpha = 0.95$, $\beta = 1.0$)						
4	1781375	33.99	1773424	34.10	$\boxed{-0.04\%}$	$\boxed{0.11}$
8	845460	31.63	838596	31.75	$\boxed{-0.08\%}$	$\boxed{0.12}$
12	505124	30.18	500135	30.29	$\boxed{-0.10\%}$	$\boxed{0.11}$
16	339817	29.21	336550	29.27	$\boxed{-0.10\%}$	$\boxed{0.06}$
20	250665	28.50	248950	28.54	$\boxed{-0.07\%}$	$\boxed{0.04}$

Table 4. Comparison of computational time of the original and the proposed quantization algorithms. Time is measured in seconds.

Sequence		$enct_o(s)$	$enct_s(s)$	$dect_o(s)$	$dect_s(s)$
Missa	(352 × 288)	20.44	20.81	0.29	0.30
Football	(352 × 288)	24.42	24.74	0.47	0.48
Boxing	(352 × 288)	24.70	25.10	0.51	0.51
Akiyo	(176 × 144)	3.00	3.05	0.08	0.08
Coastguard	(176 × 144)	5.26	5.37	0.12	0.12
River	(176 × 144)	5.90	6.01	0.12	0.12

compensation, 8% time on quantization, variable length coding and rate control, and 7% on transform coding; in decoding, around 20% is spent on transform coding, 40% on motion compensation, 25% on variable length decoding, and around 15% on dequantization.

To further understand the results in Table 4, we did an experiment to compare the time to calculate 10 000 integer operations (e.g., multiplications and divisions) and 10 000 floating point operations, respectively. We found that floating point operations result in approximately 10% increase in computational time than integer ones. This implies that if the time spent on quantization dominates the encoding procedure and the time spent on dequantization dominates the decoding procedure, then we will see 10% increase in encoding and decoding. Combined with the observation that quantization takes only 8% time in encoding and dequantization takes only 15% in decoding, it is easy to understand why the introduction of floating point quantization and dequantization in the proposed algorithms does not incur much penalty in computational time.

4. Conclusions

In this paper, we have studied reconstruction-based rate control schemes with the objective to minimize final reconstruction error when *packet losses* happen.

In general, rate control can be formulated as integer programming problems. Since it is difficult to derive signal-independent closed-form solutions to such problems, we have developed heuristic approaches to do rate control in two levels. First, for rate control among blocks within a GOB, we have studied schemes based on the "constant slope theorem", which basically states that the optimal rate allocation vector can be found at points with constant slopes in rate-distortion curves. To apply this theorem, one needs to verify an important assumption, i.e., the R-D relationship for each individual block is convex. It has been found that conventional SDC generates convex R-D curves, but no results have been reported about MDC coders with reconstructions. Our work has filled this gap by verifying empirically the convexity of R-D curves for MDC coders with reconstructions. As a result, conventional approaches based on the "constant slope theorem" can still be used for MDC coders. Second, for rate control among coefficients within a block, we have first investigated the property of coefficient variances for MDC coders. Then, based on the observations about the change of variances, we have proposed a scaled quantization scheme that produce videos with higher PSNRs using smaller bandwith.

References

1. A. Albanese, J. Bloemer and J. Edmonds, Priority encoding transmission, *Proc. Foundations of Computer Sciences* (1994), pp. 604–612.
2. R. Aravind, M. R. Civanlar and A. R. Reibman, Packet loss resilience of MPEG-2 scalable video coding algorithms, *IEEE Trans. Circuits Syst. Video Technol.* **6** (1996) 426–435.
3. V. Parthasarathy, J. W. Modestino and K. S. Vastola, Design of a transport coding scheme for high quality video over ATM networks, *IEEE Trans. Circuits Syst. Video Technol.* **7** (1997) 358–376.
4. S. D. Servetto, K. Ramchandran, V. A. Vaishampayan and K. Nahrstedt, Multiple description wavelet based image coding, *IEEE Trans. Image Processing* **9** (2000) 813–826.
5. V. A. Vaishampayan, Design of multiple description scalar quantizers, *IEEE Trans. Infor. Theor.* **39** (1993) 821–834.
6. Y. Wang, M. T. Orchard and A. R. Reibman, Multiple description image coding for noisy channels by pairing transform coefficients, *Proc. IEEE First Workshop Multimedia Signal Processing* (1997), pp. 419–424.
7. S. S. Hemami and T. H.-Y. Meng, Transform coded image reconstruction exploiting interblock correlation, *IEEE Trans. Image Processing* **4** (1995) 1023–1027.
8. S. S. Hemami and R. M. Gray, Subband coded image reconstruction for lossy packet networks, *IEEE Trans. Image Processing* **6** (1997) 523–539.
9. W. Kwok and H. Sun, Multi-directional interpolation for spatial error concealment, *IEEE Trans. Consumer Electron.* **39** (1993) 455–460.
10. J. Suh and Y. Ho, Error concealment based on directional interpolation, *IEEE Trans. Consumer Electron.* **43** (1997) 295–302.
11. W. Zeng and B. Liu, Geometric-structure-based error concealment with novel applications in block-based low-bit-rate coding, *IEEE Trans. Circuits Syst. Video Technol.* **9** (1999) 648–665.
12. M. Ghanbari, Postprocessing of late cells for packet video, *IEEE Trans. Circuits Syst. Video Technol.* **6** (1996) 669–678.
13. S. H. Lee, P. J. Lee and R. Ansari, Cell loss detection and recovery in variable rate video *Proc. 3rd Int. Workshop on Packet Video* (1990).
14. I. Rhee, Error control techniques for interactive low-bit rate video transmission over the Internet, *Proc. of ACM SIGCOMM* (1998).
15. W. Ding and B. Liu, Rate control of MPEG video coding and recording by rate-quantization modeling, *IEEE Trans. Circuits Syst. Video Technol.* **6** (1996) 12–20.
16. C.-Y. Hsu, A. Ortega and M. Khansari, Rate control for robust video transmission over burst-error wireless channels, *IEEE Trans. Circuits Syst. Video Technol.* **17** (1999) 756–773.
17. L.-J. Lin and A. Ortega, Bit-rate control using piecewise approximated rate-distortion characteristics, *IEEE Trans. Circuits Syst. Video Technol.* **8** (1998) 446–459.
18. A. Ortego and K. Ramchandran, Rate-distortion methods for image and video compression, *IEEE Signal Processing Magazine* **15** (1998) 23–50.
19. J. Ribas-Corbera and S. Lei, Rate control in DCT video coding for low-delay communications, *IEEE Trans. Circuits Syst. Video Technol.* **9** (1999) 172–185.

20. J. I. Ronda, M. Eckert, F. Jaureguizar and N. Garcia, Rate control and bit allocation for MPEG-4, *IEEE Trans. Circuits Syst. Video Technol.* **8** (1999) 1243–1258.
21. Y. Shoham and A. Gersho, Efficient bit allocation for an arbitrary set of quantizers, *IEEE Trans. Acoustics, Speech, and Signal Processing* **36** (1988) 1445–1453.
22. H. Sun, W. Kwok, M. Chien and C. H. J. Ju, MPEG coding performance improvement by jointly optimizing coding mode decisions and rate control, *IEEE Trans. Circuits Syst. Video Technol.* **7** (1997) 449–458.
23. X. Su and B. W. Wah, Streaming real-time video data with optimized reconstruction-based DCT and neural-network reconstructions, *IEEE Trans. Multimedia* **3** (2001) 123–131.
24. K. Ramchandran and M. Vetterli, Best wavelet packet bases in a rate-distortion sense, *IEEE Trans. Image Processing* **2** (1993) 160–175.
25. N. Jayant and P. Noll, *Digital Coding of Waveforms*: *Principles and Applications to Speech and Video* (Englewood Cliffs, Prentice-Hall, 1984).
26. A. Hallapuro, V. Lappalainen and T. D. Hamalainen, Performance analysis of low bit rate h.26l video encoder, *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing* (2001).
27. E. Iwata and K. Ohikotun, Exploiting coarse-grain parallelism in the MPEG-2 algorithm, Technical Report CSL-TR-98-771, Stanford University, September 1998.