MPEG Motion Picture Coding With Long-Term Constraint on Distortion Variation

Kai Wang, *Member, IEEE*, and John W. Woods, *Fellow, IEEE*

***Abstract*—A highly desirable feature in storage video applications is uniform quality. Variable bit rate (VBR) coding has the potential to produce nearly constant quality throughout an entire movie. This can be defined as a bit allocation problem with a long-term constraint on distortion variation. We consider the optimal bit allocation with multiple constraints including disk capacity and the distortion bounds on the individual frames. We find the theoretical optimality conditions and propose a practical iterative solution based on the Lagrangian methods. While minimizing average distortion and distortion variation cannot be achieved simultaneously for a given bit budget, the proposed algorithms are able to efficiently balance the tradeoff between the two goals. The computational complexity of the exact rate-distortion $(R-D)$ functions for real movies is addressed by a statistical $R-D$ model proposed in this work. Based on the generic block-based motion compensated transform coding theories, the model is formed by a rate-quantization $(R-Q)$ function and the corresponding distortion-quantization $(D-Q)$ function. A novel two-pass MPEG-2 VBR encoder based on the proposed algorithms is developed for coding with long-term nearly constant quality. Experimental results are promising and the encoder effectively achieves the fit-to-disc function and at the same time controls objective quality variation. By incorporating basic subjective coding techniques into the encoder, significant visual quality improvement is observed during the subjective tests.**

***Index Terms*—Constant quality, distortion-constrained, fit-to-disk, MPEG, perceptual coding, rate-distortion, variable bit rate (VBR).**

I. Introduction

OPTIMAL bit allocation, a classical problem in source coding, is an important open issue in digital video compression. With the evolution of techniques for recording and transmitting video signals, several international video compression standards such as MPEG-1, 2, 4, H.263, and H.264, have been established. These coding technologies, based on the common framework of block-based motion-compensated transform coding, can be used either in constant bit rate (CBR) or in variable bit rate (VBR) applications.

VBR encoding, initially used in video transmission over ATM-based broadband ISDN networks, is becoming a widely accepted coding technique in storage digital video applications. VBR encoding has the potential to produce constant visual quality throughout the entire video since the available bits can be distributed over different video segments based on their coding complexities. One important application of the VBR algorithms is a digital video disk (DVD). With a fixed disk capacity, DVD video is coded at an average bit rate of 3.5–4.0 Mbps while the peak bit rate can reach up to about 10 Mbps.

Constant quality is practically achieved by *bit allocation* of a given total bit budget through a given set of admissible quantizers. Bit allocation optimization is a fundamental issue in rate-distortion-based source coding and has been studied extensively in the literature [3]–[7]. A Lagrangian optimization approach proposed in [6] sets the foundation for much of the later work on this subject.

A typical problem encountered in video encoding applications is the *buffer-constrained* bit allocation problem which is formulated as follows.

Problem 1: Optimal buffer-constrained bit allocation is given here.

Let $\mathcal{Q}$ be a set of quantizers and let $B_{\min}$ and $B_{\max}$ be the lower and upper bounds on the bits for each source sample. Find $\mathbf{Q}^* = (Q_1^*, Q_2^*, \ldots, Q_N^*)$, with $Q_i^* \in \mathcal{Q}$ for $i = 1, 2, \ldots, N$, where N is the number of source samples, such that

$$\mathbf{Q}^* = \arg \min_{\mathbf{Q} \in \mathcal{Q}^N} \sum_{i=1}^{N} D_i(Q_i) \tag{1}$$

subject to

$$B_{\min} \le R_i(Q_i) \le B_{\max}, \quad i = 1, 2, \ldots, N \tag{2}$$

$$\sum_{i=1}^{N} R_i(Q_i) \le R_{\text{total}} \tag{3}$$

where $R_i(Q_i)$ and $D_i(Q_i)$ represent coding rate and distortion of the ith source sample using quantizer Q_i.

In addition to the total budget constraint (3), the buffer-constrained problem imposes lower and upper bounds (2) on the number of bits each source sample can produce. This is a bit allocation problem that arises in applications where the decoder buffer size is limited, and CBR coding is such an application.

The above problems use distortion as the objective function and impose constraints on rate. The roles of $D(\cdot)$ and $R(\cdot)$ in the rate-distortion function can be switched to formulate problems with the rate being the objective function. These problems usually arise in applications where constraints on distortion are desired. The *distortion-constrained* problem can be formulated as follows.

Manuscript received October 9, 2006; revised May 25, 2007. The work was supported in part by the Center for Image Processing Research and Center for Next Generation Video at Rensselaer Polytechnic Institute. This paper was recommended by Associate Editor M. van der Schaar.

K. Wang is with Qualcomm Inc., San Diego, CA 92122 USA (e-mail: kwang@qualcomm.com).

J. W. Woods is with Center for Image Processing Research, Rensselaer Polytechnic Institute, Troy, NY 12180 USA (e-mail: woods@ecse.rpi.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2008.918442

Problem 2: Optimal distortion-constrained bit allocation is given here.

Let $\mathcal{Q}$ be a set of quantizers and let $D_{\min}$ and $D_{\max}$ be the lower and upper bounds of the distortion for each source sample. Find $\mathbf{Q}^* = (Q_1^*, Q_2^*, \ldots, Q_N^*)$, with $Q_i^* \in \mathcal{Q}$ for $i = 1, 2, \ldots, N$, where N is the number of source samples, such that

$$\mathbf{Q}^* = \arg \min_{\mathbf{Q} \in \mathcal{Q}^N} \sum_{i=1}^{N} R_i(Q_i) \tag{4}$$

subject to

$$D_{\min} \leq D_i(Q_i) \leq D_{\max}, \quad i = 1, 2, \ldots, N \tag{5}$$

$$\sum_{i=1}^{N} D_i(Q_i) \leq D_{\text{total}} \tag{6}$$

where $R_i(Q_i)$ and $D_i(Q_i)$ are coding rate and distortion of the ith source sample using quantizer Q_i.

Here, the lower bound $D_{\min}$ and upper bound $D_{\max}$ specify the *distortion variation constraint* for nearly constant quality coding. If $D_{\min} = D_{\max}$, it becomes a strictly constant quality coding.

When the distortion-constrained problem is used in storage video applications, the distortion constraints are usually imposed implicitly from other requirements such as limited disk capacity. When this happens, an iterative scheme might be used to find out the equivalent distortion constraints stemmed from the application requirements.

II. Constant Objective Quality Coding

VBR is almost always used in video applications that desire nearly constant quality. With mean square error (MSE) as the distortion measure, it is called *constant objective quality*. In the literature, this is sometimes referred to as a bit allocation to *minimize distortion variation*.

It is well known that the solution to Problem 1 can be found by Lagrangian optimization ([6], [8], [9]). To be specific, the problem can be solved by solving an equivalent unconstrained problem with the objective function defined as

$$J_i(\lambda, Q_i) \triangleq D_i(Q_i) + \lambda R_i(Q_i) \tag{7}$$

where λ is a nonnegative real number called the *Lagrangian parameter*.

The Lagrangian optimization is sometimes referred to as a constant slope solution in the sense that it implies the optimal operating point is located where the slopes of the $R-D$ curves of the individual coding units are the same. The Lagrangian parameter λ being admissible essentially means the operating points must stay within the "window" between $B_{\min}$ and $B_{\max}$. Finally, due to the discreteness of the Q values in standard video codecs, λ cannot be any arbitrary real value and must correspond to admissible Q values. This is illustrated in Fig. 1(a) with a clip consisting of three frames.

We assume the experimental $R-D$ curve is convex, as would be the theoretical curve. The constant slope approach also applies to the distortion-constrained problem. By exchanging $D(\cdot)$ and $R(\cdot)$ in $R-D$ function and constraint equations, we have a corresponding distortion-constrained problem as illustrated in Fig. 1(b). However, an important difference between the buffer-constrained coding and distortion-constrained coding is that the distortion constraints $D_{\min}$ and $D_{\max}$ are usually fixed values set before the encoder is launched. This is because the distortion-constrained problem usually arises in applications where the variation of the video quality must be confined within a certain range. In this work we use D_T, called *target distortion*, to denote the center of the constraint window

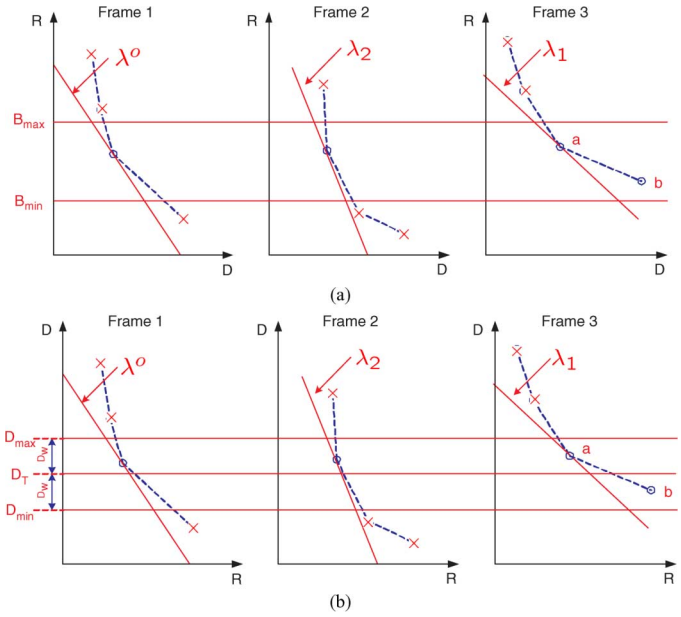

Fig. 1. Constant slope solution budget-constrained and distortion-constrained problems. (a) Buffer-constrained bit allocation, $B_{\max} > B_{\min} > 0$. (b) Distortion-constrained bit allocation.

$$D_T \triangleq \frac{D_{\max} + D_{\min}}{2} \tag{8}$$

and introduce a *tolerance* α to control the window size depending on the target distortion

$$\alpha \triangleq \frac{D_{\max} - D_{\min}}{2D_T}. \tag{9}$$

We then can express the constraints in Problem 2 in terms of D_T and α as

$$(1-\alpha)D_T \leq D_i(Q_i) \leq (1+\alpha)D_T \tag{10}$$

$$\sum_{i=1}^{N} D_i(Q_i) \leq N \times D_T \tag{11}$$

where N is the total number of frames and $N \times D_T$ is equal to D_{total} in (6).

Similar to the Lagrangian approach used for the buffer-constrained problem in [8], we first propose an iterative approach to solve the distortion-constrained problem for independent coding, e.g., MPEG I-frame-only coding, with the distortion constraints (10) and (11).

Algorithm I: *Optimal Distortion-Constrained Independent Coding*

1) Consider this as an equivalent problem with a total "distortion budget" $D_{\text{total}} = N \times D_T$.
2) Apply the Lagrangian method to solve the special case of **Problem 2** with $D_{\min} = 0$ and $D_{\max} = \infty$. And the result is the constant slope solution with optimal λ^o and corresponding $\mathbf{Q}^o$. It is known from the convex optimization [25] that the optimality is achieved when
$$\sum_{i=1}^{N} D_i(Q_i) \approx D_{\text{total}}. \tag{12}$$
The approximation is due to the fact that the operational rate-distortion function is a discrete function.
3) Impose the distortion constraints.
The quantization found in the previous step is the optimal solution that minimizes total bit rate for a given total distortion. For any frame i, the constraint condition $D_{\min} \le D_i(Q_i^o) \le D_{\max}$ may be violated. Depending on the value of $D_i(Q_i^o)$, frames are divided into three groups:
   a) If $D_i(Q_i^o) < D_{\min}$, the constant slope solution for this frame is not admissible. In this case, we need to replace Q_i^o by an admissible Q_i^* such that $D_i(Q_i^*) \approx D_{\min}$.
   b) Similarly, if $D_i(Q_i^o) > D_{\max}$, replace Q_i^o by an admissible Q_i^* so that $D_i(Q_i^*) \approx D_{\max}$.
   c) If $D_{\min} \le D_i(Q_i^o) \le D_{\max}$, the constant slope solution $D_i(Q_i^o)$ does not violate the distortion constraints, but, due to the changes of the operating points of the frames in the other two groups, the Q values for this group cannot be finalized at this stage.
4) Initialize the next iteration. Let the number of frames in the above three groups be $N_{\min}$, $N_{\max}$, and N_{mid}, respectively. If $N_{\min} = N_{\max} = 0$, then the constant slope solution found in Step 2) is also the solution to the given distortion-constrained problem. Otherwise, perform the following substitutions:
$$N \longleftarrow N_{mid}, \quad \text{and} \tag{13}$$
$$D_{\text{total}} \longleftarrow D_{\text{total}} - N_{\min}D_{\min} - N_{\max}D_{\max}. \tag{14}$$
If the updated D_{total} value is positive, go back to Step 2); otherwise, end the algorithm and the $\mathbf{Q}$ found at this iteration is the approximated solution to the given distortion-constrained problem.

During each iteration, the proposed algorithm starts with the constant slope solution to a distortion-constrained problem with an updated D_{total} from the previous iteration. It then moves the operating points of the frames that violate the additional distortion constraints to the boundary $D_{\min}$ or $D_{\max}$. The iteration continues until no more frames violate the constraints. This algorithm produces a globally optimal solution for the given distortion-constrained problem. It is optimal in the sense that it is the best possible result for the given constraints. The optimality of this algorithm is justified in [23] using the Karush–Kuhn–Tucker condition. Interested readers can refer to [25] for background on convex optimization.

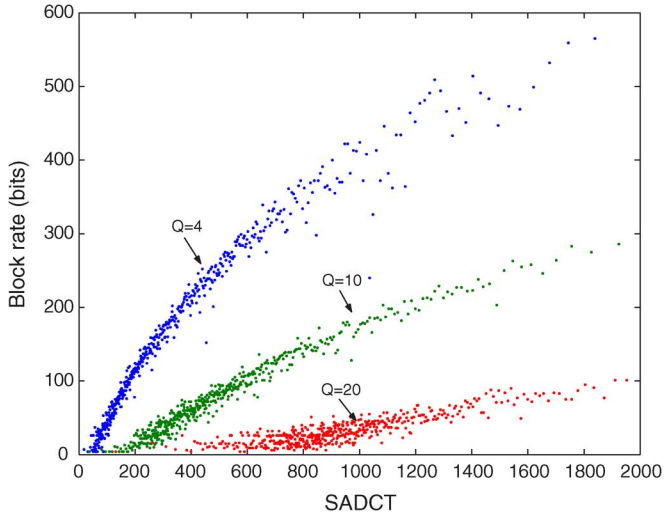

Fig. 2. Average block bits $r_b = r_b(Q, \mathbf{S})$ with $\mathbf{S} = \text{SADCT}$. Data are generated from *Flower garden*.

III. Novel Two-Pass Encoder

A two-pass encoder is proposed in this study. In the first pass, the frame-level $R - D$ characteristics are estimated in terms of $R - Q$ and $D - Q$ functions. Also, in the second pass, the estimated functions are used by the proposed nearly constant distortion bit allocation algorithms. These algorithms are also able to consider the trade-off of minimizing distortion variation and total distortion while achieving the fit-to-disk function given the target disk capacity. For more accurate motion estimation and $R - D$ modeling, the first pass is encoded at a higher quality level than the second pass is. The mode decision and motion information obtained in the first pass will be reused in the second pass.

A. Bit Rate Estimation

Many practical $R - Q$ models have been developed in the literature [10]–[14] to address the problem that the theoretical Gaussian or Laplacian models are not accurate enough in real coding. In this study, we propose a statistical DCT block level $R - Q$ model that can be used in a two-pass encoder to estimate the DCT AC rate. The model can be represented by
$$r_b = r_b(Q, \mathbf{S}) \tag{15}$$
where Q is the quantization parameter and $\mathbf{S}$ is the block statistic. After extensive experimental comparison, we find the *sum of absolute DCT coefficients* (SADCT), defined as
$$\mathbf{S} \triangleq \sum_{k=k_0}^{64} |C_k| \tag{16}$$
where C_k is the kth DCT coefficient, to be a good statistic for the block level rate estimate. Here, the start index k_0 indicates if the dc component of a transform block is counted for SADCT. The value of k_0 is 1 for intra macorblock and 0 for inter macroblock. Fig. 2 shows such an example.

The model is trained before the first pass using selected video clips and stored as lookup tables (LUTs). With the proposed two-pass encoder where the first-pass mode decision and motion information are reused in the second pass, we can simply look

up the $R-Q$ tables to estimate the intra-block rate since the intra-block SADCT values are the same in both passes. However for inter-blocks, we have to deal with the issue that their SADCT values are different because their reference blocks are likely to be quantized using a different Q in the second pass. A correction factor $\hat{\gamma}$, a function of the reference frame quantization parameters $Q_1^{(r)}$ for the first pass and $Q_2^{(r)}$ for the second pass, is introduced to compensate for this discrepancy. As a result, the estimate for the second pass $\widehat{\mathrm{SADCT}}_2$ can be obtained by scaling the actual value of the first pass SADCT_1 as

$$\widehat{\mathrm{SADCT}}_2 = \hat{\gamma}\left(Q_2^{(r)}, Q_1^{(r)}\right) \times \mathrm{SADCT}_1. \tag{17}$$

The details on how to estimate $\hat{\gamma}$ can be found in Appendix II.

For dc and overhead bits, we estimate separately. Considering the coding quality of the second pass shall not be far different from first pass, we use the same dc quantization parameter in both passes and thus we can directly use the first-pass dc bit rate as an estimate. For overhead bits, since we reuse the mode decisions, the main difference between the two passes is contributed by the extra number of macroblocks that may be skipped in the second pass due to the increase of Q implied by the fact that the first pass is encoded at a higher quality than the application requires. This can be estimated from Q_2 and the maximum DCT coefficient in each block.

Fig. 3 shows the simulation results of the relative error in estimating the second-pass bit rate. The relative error is defined as

$$\Delta_r \triangleq \frac{\bar{R}_{\mathrm{est}} - \bar{R}_{\mathrm{act}}}{\bar{R}_{\mathrm{act}}} \tag{18}$$

where $\bar{R}_{\mathrm{act}}$ and $\bar{R}_{\mathrm{est}}$ are the actual and estimated average bit rates, respectively. For comparison, the estimation error of the simple method without the correction (17) is also shown.

B. DCT Distortion Estimation

To obtain a $D-Q$ estimate we introduce a computational $D-Q$ model based on previous work [15] that originated from the distortion-rate function for the memoryless Gaussian source. Such a model is motivated by the fact that the well known *fine quantization approximation (FQA)* is not accurate enough for real data. The model in [15] indicated that the MSE of a DCT coefficient can be estimated by

$$D_k = \begin{cases} \frac{\Delta^2}{\beta}, & \text{if } \sigma_k \geq T_k \\ \sigma_k^2, & \text{if } \sigma_k < T_k \end{cases} \tag{19}$$

where T_k is a threshold depending on the relationship between the quantization step size Δ and the signal standard deviation σ_k. Unlike FQA, where β is always 12, in this model β is a nonlinear function of Δ/σ_k

$$\eta \triangleq \ln\left(\frac{\Delta}{\sigma}\right) \tag{20}$$

$$\beta = \beta(\eta). \tag{21}$$

In order to use this model in real coding, the $\beta-\eta$ function must be found by training.

The model does not take into account the fact that for sources with different statistics, the $\beta(\eta)$ functions can be very different.

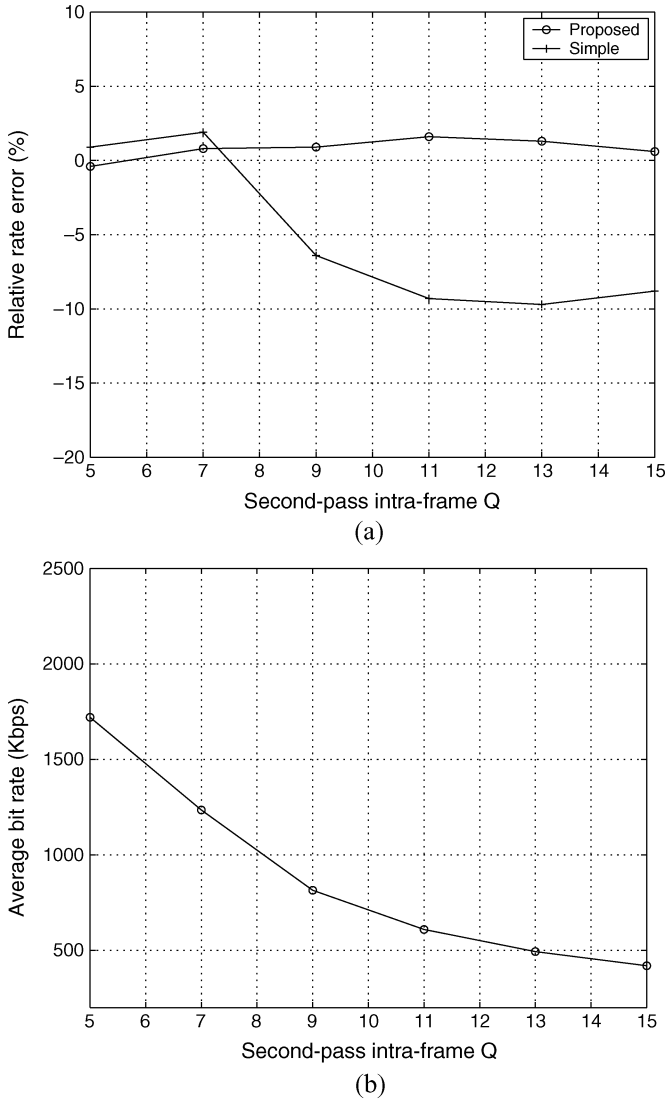

Fig. 3. Rate estimate for SIF movie sequence *Mclip2* (86 400 frames) using proposed algorithms. Constant-Q encoder is used in both passes. First-pass $\mathbf{Q_1} = (Q_I, Q_P, Q_B) = (7, 8, 8)$. Second-pass $\mathbf{Q}_2 = (Q_I, Q_P, Q_B) = (Q_I, Q_I + 1, Q_I + 1)$ with $Q_I = Q$ given in the figure. *Table Tennis* is used as the training sequence.

In order to address this issue, we extend this model by introducing a *model mismatch parameter* ε as

$$D_k = \begin{cases} \frac{\Delta^2}{\varepsilon\beta}, & \text{if } \eta \leq \eta_T \\ \sigma_k^2, & \text{if } \eta > \eta_T \end{cases} \tag{22}$$

where the threshold conditions are now expressed in terms of η, the log of the ratio of Δ and σ.

This model can be used in a two-pass encoder to achieve higher distortion estimation accuracy. Since the same distortion model is assumed for both the training and the test data, the first pass result gives certain information on how closely the test data match the training data. And we can obtain an estimate of the parameter ε from the first pass result using

$$\varepsilon = \frac{D_{k,T}(Q_1)}{D_k(Q_1)}\varepsilon_T \tag{23}$$

where $D_{k,T}(Q_1)$ and $D_k(Q_1)$ are the actual distortion values of the training data and the test data at Q_1. The model parameter ε_T

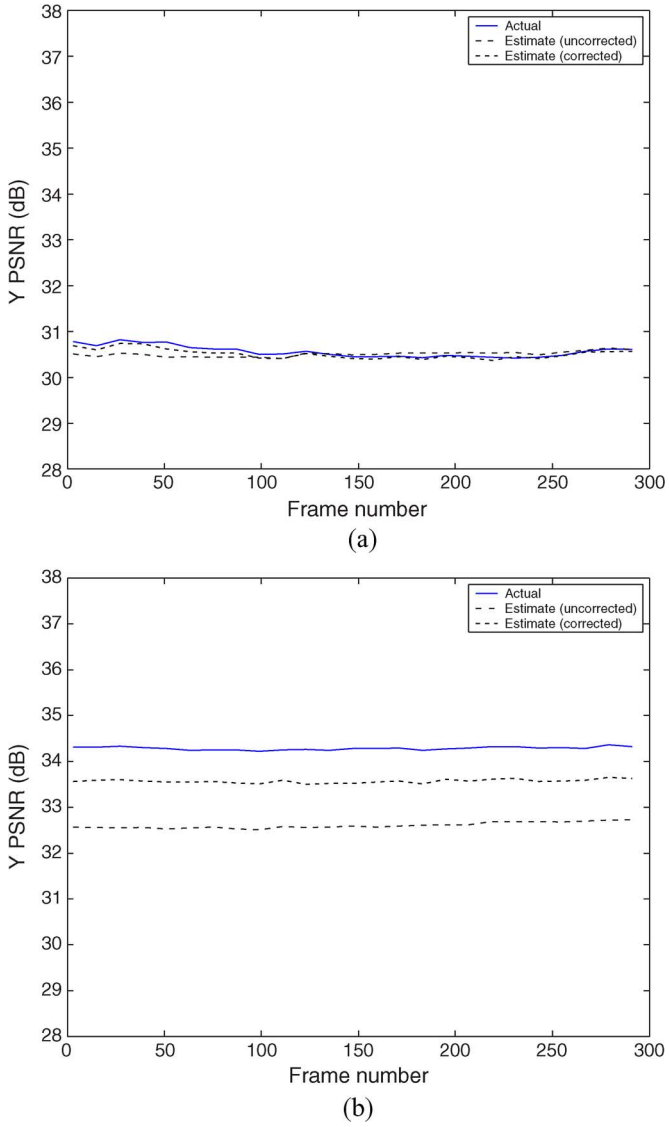

Fig. 4. I-frame PSNR estimate using training sequence *Coastguard*. Intra-frame quantization parameter $Q_1 = 7$ in the first pass and $Q_2 = 13$ in the second pass. (a) Mobile. (b) Container.

is for the training data and is obtained together with other model parameters during the training process. If the test data matched the training data perfectly, ε would be equal to ε_T. The experimental results in Fig. 4 show that the above mismatch correction can be very effective in improving the PSNR estimation.

For inter-blocks, similar to the bit rate estimation, we have to compensate for the discrepancies between the first- and second-pass DCT variances. Based on the derivation in Appendix I, the block variance σ_k^2 used in (19)–(22) shall be the compensated value

$$\sigma_k^2 \leftarrow \sigma_k^2 \left(1 + \frac{\bar{D}_k^{(r)}\left(Q_2^{(r)}\right) - \bar{D}_k^{(r)}\left(Q_1^{(r)}\right)}{\sigma_k^2}\right) \tag{24}$$

where $\bar{D}_k^{(r)}(Q_2^{(r)})$ and $\bar{D}_k^{(r)}(Q_2^{(r)})$ are the same as in (43). However, since the second-pass $\bar{D}_k^{(r)}(Q_2^{(r)})$ is unknown until the reference block is actually encoded, we have to first estimate $\bar{D}_k^{(r)}(Q_2^{(r)})$ in order to get a compensated σ_k^2. Due to this, the distortion estimation must be performed in encoding order such that the predictor distortion $\bar{D}_k^{(r)}(Q_2^{(r)})$ is always estimated before the current block.

TABLE I
RELATIVE ERROR IN P-FRAME DISTORTION ESTIMATION WITH AND WITHOUT MODEL MISMATCH CONTROL (MMC). THE FIRST-PASS QUANTIZATION PARAMETER $Q_1 = 7$. TEST VIDEO *Mclip1* IS A SHORT MOVIE CLIP

Training clip		*Flower Garden*		*Table Tennis*		*Coastguard*	
Test clip	Q_2	mmc off	mmc on	mmc off	mmc on	mmc off	mmc on
Container	11	34%	13%	39%	17%	35%	16%
	13	46%	16%	50%	21%	48%	19%
	15	55%	18%	58%	22%	56%	21%
Mobile	11	18%	12%	19%	4%	9%	9%
	13	20%	18%	22%	6%	13%	12%
	15	22%	23%	26%	7%	18%	15%
Silent	11	39%	18%	42%	17%	39%	17%
	13	47%	21%	50%	19%	47%	19%
	15	53%	22%	56%	20%	54%	21%
Mclip1	11	12%	9%	12%	11%	11%	8%
	13	14%	10%	16%	13%	14%	9%
	15	18%	11%	20%	15%	16%	10%

Table I shows the simulation results of the relative error in estimating the second-pass distortion. Similar to bit rate estimation, the relative error defined for distortion estimation is defined as

$$\Delta_d \triangleq \frac{|\bar{D}_{\text{est}} - \bar{D}_{\text{act}}|}{\bar{D}_{\text{act}}} \tag{25}$$

where $\bar{D}_{\text{act}}$ and $\bar{D}_{\text{est}}$ are the actual and estimated average distortion in terms of MSE, respectively. The simulation results verify that the mismatch correction is useful for different training clips and test clips as well as different second pass quantization parameters.

C. Quantizer Estimation

The quantizer estimation is the core decision part of the proposed two=pass encoder. After the first pass yields the $R - Q$ and $D - Q$ estimates based on the proposed models, the objective of the quantizer estimation is to find the frame-level quantization parameters $\mathbf{Q}^{(2)} = (Q_1^{(2)}, Q_2^{(2)}, \ldots, Q_N^{(2)})$ that can be used by the second pass to produce a nearly constant distortion bitstream. If we denote the estimation by $\mathbb{G}$, it consists of two components, a distortion-to-quantization estimation $(\mathbb{G}_{D\to Q})$ and a quantization-to-rate estimation $(\mathbb{G}_{Q\to R})$. They are used in an iterative procedure to find the $\mathbf{Q}^{(2)}$ for the given disk capacity C_{tot} and *target fullness* κ. The algorithm is summarized as follows and illustrated in Fig. 5.

1) Specify an initial value for the target distortion D_T. This value must be small enough so that if the second pass operates at this distortion level the output bitstream will overflow the disk. This is to ensure search for the final operating points proceeds in one direction.
2) Perform $\mathbb{G}_{D\to Q}$ using the given D_T and a predetermined tolerance α as inputs to get a *candidate quantizer* $\mathbf{Q}^*$.
3) Perform $\mathbb{G}_{Q\to R}$ using the above $\mathbf{Q}^*$ as input to obtain an estimate of the total bits $\hat{R}_{\text{tot}}$. Since we already have the

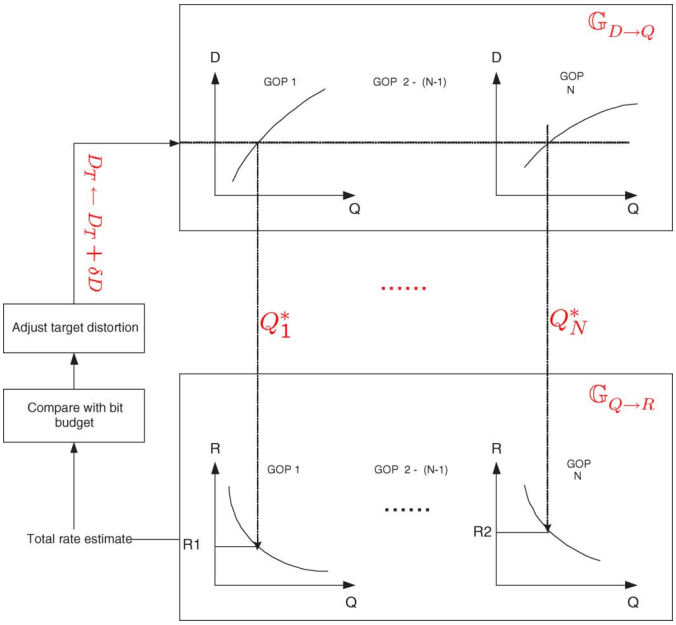

Fig. 5. Schematic for the proposed iterative quantizer estimation.

frame-level $R-Q$ estimate from the first pass, this step is simply an LUT for each frame.

4) Compare $\hat{R}_{\text{tot}}$ with κC_{tot}. If $\hat{R}_{\text{tot}} > \kappa C_{\text{tot}}$, let $D_T = D_T + \delta D$ where $\delta D > 0$ and go to Step 2); otherwise, end the iteration and $\mathbf{Q}^{(2)} = \mathbf{Q}^*$ is the final solution.

During each iteration, the $\mathbb{G}_{D\to Q}$ step tries to find the optimal solution to the distortion-constrained bit allocation problem specified by the given D_T and α. While the given D_T might not lead to a solution that meets the target disk capacity, the iterative procedure is a scheme that is able to search and eventually find such a D_T for the final solution.

The $\mathbb{G}_{D\to Q}$ step is where the algorithm for optimal distortion-constrained bit allocation problem is implemented. Considering both the Lagrangian theory and the practical issues, a nearly constant distortion algorithm is proposed in the next section.

D. Nearly Constant Distortion Algorithm

The nearly constant distortion coding requires the distortion of each frame to stay within the distortion constraint jointly specified by D_T and α defined in (8) and (9). In addition, we consider the *monotonicity* property [22], which indicates that, for predictive coding, a better predictor will lead to more efficient coding. In order to use this property to preserve the overall coding efficiency, in each group of pictures (GOP), we restrict the quantization parameters of P- and B-frames to be no smaller than those of the I-frame, using ΔQ_P and ΔQ_B to represent the difference. The proposed algorithm is described as follows.

Algorithm II: *Nearly Constant Distortion Coding*

1) I-frame quantization estimation
For the given target distortion D_T and tolerance α, apply **Algorithm I** to find the quantization parameter $\mathbf{Q}_I^*$ for the I-frames. $\hat{\mathbf{D}}_I(\mathbf{Q}_I^*)$ is then available too.
2) Preliminary P-, B-frame quantization estimation
For each P- or B-frame, apply the proposed $D-Q$ model to look up the Q values that will produce D values closest to the corresponding I-frame distortion D_I.
3) Refine P-, B-frame quantization estimation Within each GOP, if the preliminary estimation for a P-frame Q_P^* is smaller than $Q_I^* + \Delta Q_P$, where Q_I^* is the corresponding I-frame quantization parameter determined in Step 1, set $Q_P^* = Q_I + \Delta Q_P$. For B-frames, if Q_B^* is smaller than $(Q_{\text{fwd}}^* + Q_{\text{bwd}}^*) + \Delta Q_B$, where Q_{fwd}^* and Q_{bwd}^* are the quantization parameters for the corresponding forward and backward reference frames, set $Q_B = (Q_{\text{fwd}}^* + Q_{\text{bwd}}^*) + \Delta Q_B$.
4) With the above two steps, $\mathbf{Q}^{(2)} = \mathbf{Q}^*$ is the output.

This algorithm provides a way to use the tolerance variable α and monotonicity parameters ΔQ_P and ΔQ_P to balance the tradeoff between minimizing total distortion and distortion variation for a given C_{tot}.

The monotonicity property holds only if the temporal prediction is good enough such that most of the macroblocks in the current frame can find matching macroblocks in the reference frame(s). This is only possible when there is no scene change between the current frame and the reference frame(s). In order for **Algorithm II** to produce a near optimal solution, both scene change detection and dynamic GOP size are implemented in our encoder.

IV. Experimental Results

In our simulations, the proposed algorithms are implemented in a two-pass MPEG-2 encoder. The test clips used in the experiments are mainly SIF clips digitized from DVD quality standard resolution movies. We also use a real movie source *Stavro (To Kill a Lawyer)* at the resolution 560 × 320. All of the SIF clips have 86 400 frames and *Stavro* has 123 840 frames, all coded at 24 frames/s.

Fig. 6 shows the simulation results of **Algorithm II** using one of the SIF clips. The $D(R)$ curves in Fig. 6(a) show that the average coding efficiency of the proposed two-pass VBR can be adjusted using the tolerance α. A larger α means more distortion variation is allowed and at the same time better compression can be achieved. Correspondingly, Fig. 6(b) shows that the PSNR variance as a measure of distortion variation is a function of α. The figure does not show the PSNR variance of the TM5 encoder because its value is in another order of magnitude (≈20) and may hide the details of these curves if plotted together. Similar results were observed in all the experiments. This clearly shows the algorithm is able to balance the tradeoff between total distortion and distortion variation.

We show next that the two-pass VBR encoder can achieve the "fit-to-disk" function while preserving the nearly constant quality. Each test clip is encoded with two different virtual disk sizes, targeting for high and medium bit rates. Tables II and III show that the encoder successfully generates bitstreams that are all able to fit the target disks with a fullness upwards of 90%. In all the experiments, the PSNR variance of the proposed two-pass VBR encoder is significantly smaller than that of the TM5 CBR and constant-Q encoders running at approximately

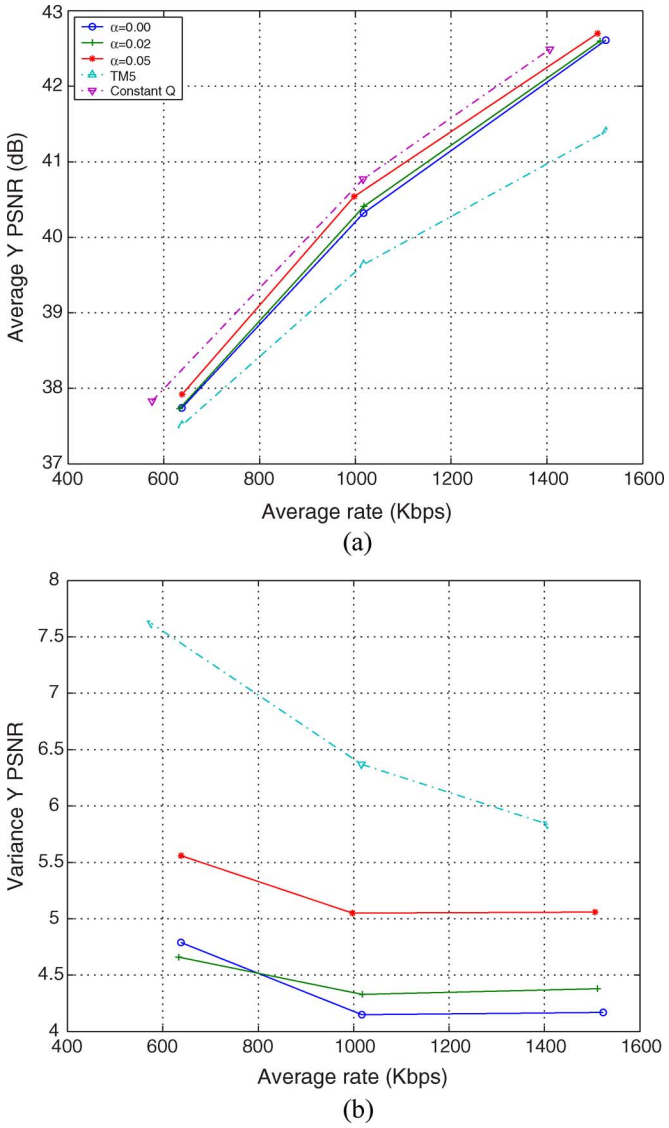

Fig. 6. Nearly-constant-distortion **Algorithm II** simulation results. Three different constraint window parameters are compared. TM5 CBR coding and Constant-Q coding results are also shown. Constant-Q coder uses $\mathbf{Q} = (Q_I, Q_P, Q_B) = (Q_I, Q_I + 1, Q_I + 1)$ and different Q_I values are tried to match the rates of other coders. (a) *Mclip5a*, average PSNR. (b) *Mclip5a*, variance PSNR.

the same average bit rates. Table IV compares in detail the differences among the three encoders with the SIF clip *Mclip3d*, and Fig. 7 shows the corresponding PSNR plots.

V. Towards Constant Subjective Quality

We have so far developed a two-pass encoder that is able to achieve nearly constant objective quality for a given disk capacity. The encoder has shown success with a large variety of source clips at medium-to-high bit rates. While the high-rate output bitstreams are nearly visually lossless with an almost constant PSNR, the nonuniform visual distortion is noticeable and becomes increasingly annoying as the bit rate decreases. This is not surprising because the *human visual system* (HVS) is known not to favor PSNR. So, here, and based on our objective two-pass VBR encoder, we incorporate three subjective coding techniques to move towards perceptually constant quality for the medium bit rate case.

TABLE II
Simulation Results of **Algorithm II** ($\alpha = 0.02$) With Large-Size Virtual Disks

	PSNR		Rate	Virtual disk (target fullness: 97.0%)			
Clips	Mean (dB)	Var.	$\bar{R}$ (Kbps)	$\hat{R}_{tot}$ (MBytes)	R_{tot} (MBytes)	C_{tot} (MBytes)	Fullness
Mclip1a	42.80	4.75	1473.0	701.0	662.6	720	92.1%
Mclip2a	40.86	6.15	1498.6	700.9	674.4	720	93.7%
Mclip3a	38.93	0.93	1486.1	701.6	668.8	720	92.9%
Mclip4a	43.01	3.32	1485.8	701.2	668.6	720	92.9%
Mclip5a	42.69	4.52	1477.6	701.9	664.9	720	92.4%
Mclip6a	44.41	4.88	1475.2	701.0	663.8	720	92.2%
Mclip7a	42.44	2.45	1471.9	699.5	662.3	720	92.0%
Mclip1b	43.64	0.76	1477.3	699.6	664.8	720	92.3%
Mclip2b	41.40	0.51	1471.3	700.2	662.1	720	92.0%
Mclip3b	39.57	0.65	1485.5	701.3	668.5	720	92.8%
Mclip4b	43.14	1.21	1480.2	701.6	666.1	720	92.5%
Mclip5b	42.34	4.66	1480.8	701.0	666.4	720	92.6%
Mclip6b	44.60	1.41	1482.2	700.5	667.0	720	92.6%
Mclip7b	42.10	1.62	1475.6	699.3	664.0	720	92.2%
Stavro	40.95	4.62	3203.4	2045.0	2066.2	2100	98.4%

TABLE III
Simulation Results of **Algorithm II** ($\alpha = 0.02$) With Medium-Size Virtual Disks

	PSNR		Rate	Virtual disk (target fullness: 97.0%)			
Clips	Mean (dB)	Var.	$\bar{R}$ (Kbps)	$\hat{R}_{tot}$ (MBytes)	R_{tot} (MBytes)	C_{tot} (MBytes)	Fullness
Mclip1a	40.40	5.02	994.7	467.9	447.6	480	93.3%
Mclip2a	38.50	6.48	1012.7	467.9	455.7	480	94.9%
Mclip3a	36.42	0.98	978.8	466.3	440.5	480	91.8%
Mclip4a	40.66	3.14	1003.8	467.9	451.7	480	94.1%
Mclip5a	40.54	4.42	993.6	467.9	447.1	480	93.2%
Mclip6a	42.59	4.61	964.7	466.0	434.1	480	90.4%
Mclip7a	40.00	2.53	975.8	455.6	439.1	480	91.5%
Mclip1b	41.27	0.66	974.9	466.3	438.7	480	91.4%
Mclip2b	38.97	0.47	972.6	467.2	437.7	480	91.2%
Mclip3b	37.23	0.81	984.1	467.2	442.8	480	92.3%
Mclip4b	40.90	1.30	993.3	418.2	447.0	480	93.1%
Mclip5b	39.98	4.57	980.8	457.0	441.4	480	92.0%
Mclip6b	42.83	1.30	973.3	467.6	438.0	480	91.2%
Mclip7b	39.67	1.78	1001.2	465.7	450.6	480	93.9%
Stavro	39.70	4.61	1001.2	1362.7	1272.6	1400	90.8%

1) **Weighted PSNR**

The *Contrast sensitivity function (CSF)* implies different sensitivity of the HVS to different spatial frequencies. We adopt CSF-weighted MSE as a distortion measure. The weighted MSE (WMSE) is calculated in the DCT domain

$$\mathbf{WMSE}^{(\mathrm{CSF})} \triangleq \frac{1}{64}\sum_{k=1}^{64} W_k D_k \tag{26}$$

where D_k is MSE of the kth coefficient and W_k is the corresponding weighting factor. As the equation shows,

TABLE IV
COMPARE TM5 CBR, CONSTANT-Q AND THE PROPOSED TWO-PASS VBR RESULTS

		PSNR		Rate	Frame *mquant* (I P B)	
Clips	Enc.	Mean (dB)	Var.	$\bar{R}$ (Kbps)	Mean	Var.
High bit rate, large disk size						
Mclip3d	VBR	42.21	0.89	2366.6	(4.8,5.9,6.7)	(1.6,1.0, 1.1)
	TM5	40.74	15.22	2398.4	(4.9,5.0,7.0)	(6.6, 4.4, 8.9)
	Const-Q	42.29	1.50	2376.3	(5.0,6.0,6.0)	(0.0, 0.0, 0.0)
Medium rate, medium disk size						
Mclip3d	VBR	37.68	0.97	1185.5	(12.2,13.4,15.2)	(5.0,4.0,4.1)
	TM5	37.11	21.2	1207.2	(9.1,9.6,13.4)	(22.5,21.2,45.0)
	Const-Q	38.23	4.38	1220.3	(10.0,12.0,12.0)	(0.0,0.0,0.0)

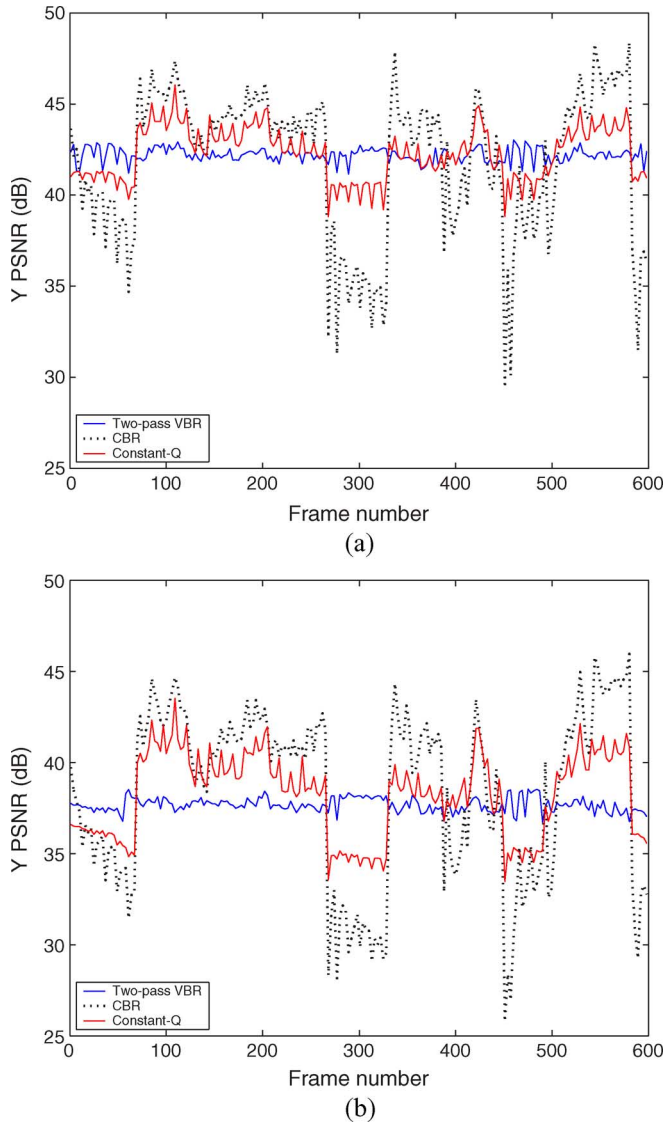

Fig. 7. PSNR of *Mclip3d* using **Algorithm II** ($\alpha = 0.02$) at (a) high and (b) medium bit rates.

WMSE must be calculated on the DCT domain and in order for the **WMSE** to be in the same numerical scale as the regular MSE, the weighting matrix must be normalized as

$$\sum_{k=1}^{64} W_k = 1. \tag{27}$$

We compared different CSF weighting matrices in the literature [1], [2], [17]–[19] and find the MPEG-2 default weighting matrix is a reasonably good candidate.

2) **Relative *global complexity (RGC)***
By visually inspecting the quality of the encoded video clip, we noticed that distortion is more obvious in frames with large flat area. To obtain better visual quality, we need a biased bit allocation in favor of those low detail frames. This is achieved by employing a global complexity weighted MSE, defined as

$$\mathbf{WMSE}^{(\mathrm{RGC})} = \frac{\overline{\breve{\sigma}^2}}{\breve{\sigma}^2} D_f \tag{28}$$

where $\breve{\sigma}^2$ is the *global variance*, defined in [20] as the mean of the block variance across the entire frame. The variable $\overline{\breve{\sigma}^2}$ is the mean of $\breve{\sigma}^2$ over the entire clip.

3) **Spatial visual masking by macroblock classification**
Within each frame, we further explore the spatial visual masking by assigning more bits to macroblocks containing low detail texture and edges. In order to do this, we first classify a macroblock into either a texture macroblock or an edge macroblock based on the distribution of the block-level complexity in terms of block variance. Then for either type, we assign a weighting factor to each macroblock. The weighting ensures that low detailed texture macroblocks are assigned more bits than busy ones and edge macroblocks are favored to some extent during bit allocation. The weighting factors typically depend on bit rates and are obtained by extensive subjective evaluation. During the encoding, the macroblock level quantization parameters are adjust appropriately in order to achieve more uniform visual quality within a frame.

The above three techniques are applied at different stages during the encoding. The RGC is used to adjust the frame-level rate control and the spatial masking is used to adjust the macroblock-level rate control. The weighted PSNR replaces the regular PSNR for distortion estimation so that our constant-distortion algorithms become constant weighted distortion.

To verify that these methods improve the visual quality, we ran a simple subjective test using a CIF clip consisting of six concatenated standard MPEG test clips. We encode this composite clip using three different encoders with approximately the same average bit rate. During the subjective test, we showed two clips side-by-side each time in a properly darkened room with the viewers at the appropriate distance from a high-quality CRT monitor. We first asked the viewers if they notice any difference between the two clips. If no quality difference is observed, both clips get scored as 0, otherwise they were asked to provide a score to the one they considered better, with 1 for significantly better and 0.5 for slightly better. The overall subjective test results are given in Table V. It is worth mentioning that the preference score is only an indication of which clips are

TABLE V
SUBJECTIVE TESTS COMPARING THE PROPOSED SUBJECTIVE VBR, TM5 CBR, AND CONSTANT-Q VBR ENCODERS. THE SIX TEST CLIPS HAVE DIFFERENT SPATIAL AND TEMPORAL COMPLEXITIES. THE SCORE IS BASED ON TEN VIEWERS

	Bit rate (Kbps)			Sum preference score			
Clips (frames)	VBR	CBR	Const-Q	VBR	CBR	VBR	Const-Q
Container (300)	502.3	1021.1	344.9	0.0	5.0	9.0	0.0
Mobile (300)	1520.1	1021.1	1665.9	6.0	0.0	0.0	0.0
Silent (300)	708.6	1021.1	355.5	0.0	1.0	9.0	0.0
Stefan (300)	1583.5	1021.3	1880.0	9.5	0.0	1.5	0.0
Table Tennis (300)	895.7	1020.8	520.7	1.0	2.5	0.0	0.0
Bus (150)	965.5	1020.8	1581.7	9.0	0.0	0.0	0.0
Overall	1035.1	1021.1	1010.5	25.5	8.5	19.5	0.0

preferred and how much they are preferred. It is not a quantitative measure for the visual quality. As a result, the same value of the preference score in different experiments does not imply the same visual quality. Among the six selected clips, we consider *Container*, *Silent*, and *Table Tennis* to be low-activity clips. The results show that, between the CBR and the subjective VBR, CBR results in slightly better visual quality for low-detail clips while the subjective VBR is considered to be much more visually appealing, for the active clips. And compared with the constant-Q VBR, the proposed subjective VBR is always voted as preferred when a preference is noted.

VI. CONCLUSION

In this study, we developed a nearly constant quality coding algorithm using a two-pass VBR encoder. The problem was posed as distortion-constrained bit allocation that can be solved using a Lagrangian method. We also developed statistical rate-distortion models that are used by the proposed two-pass encoder to resolve computational complexity of the exact rate-distortion functions for real movies. We have shown that the proposed algorithms can trade off between minimizing average distortion and distortion variation and at the same time achieve the fit-to-disk function for a given disk capacity. We also incorporate some basic subjective coding techniques to improve the visual quality for medium bit rate applications.

Since our algorithms are based on the common framework of block-based motion-compensated DCT transform coding, they can potentially be extended to other coding standards including H.264. A further improvement could be a more elaborate subjective coding technique to achieve improved perceptual quality.

APPENDIX I
INTER-BLOCK RESIDUAL STATISTIC COMPENSATION IN TWO-PASS ENCODER

DCT block variance is an important statistic used throughout this work for both bit-rate and distortion estimation. For inter-blocks, the second-pass residual block variance is different from the first pass value because the quantization of the reference block differs from the first pass due to bit allocation. To obtain a more accurate $R-D$ estimate for the second pass, we have to compensate this discrepancy when the first-pass statistics including the block variance are used during the estimation. Also, the compensation factor can be developed from DPCM coding theory and uniform quantization properties.

The block-based motion-compensated predictive coder is essentially a closed-loop DPCM coder. With such a DPCM coder, the quantizer input (y) is the *prediction error*

$$y = x - \hat{x} \tag{29}$$

where $\hat{x}$ is a prediction of the source signal x. With block-based motion-compensated coders, the above signals are all motion-compensated blocks with the prediction also called the *reference block*. Being closed-loop implies that the reference block is the same as the decoder-reconstructed block $\tilde{\mathbf{x}}^{(r)}$

$$\mathbf{y} = \mathbf{x} - \tilde{\mathbf{x}}^{(\mathbf{r})}. \tag{30}$$

Since DCT is a linear operation, it also applies in the DCT domain as

$$\mathbf{Y} = \mathbf{X} - \tilde{\mathbf{X}}^{(\mathbf{r})} \tag{31}$$

where $\mathbf{Y}$, $\mathbf{X}$, and $\mathbf{X}^{(r)}$ are the corresponding blocks of DCT coefficients. For the individual DCT coefficients, we have

$$Y_k = X_k - \tilde{X}_k^{(r)}. \tag{32}$$

We can further express Y_k in terms of the *open-loop prediction error* $Y_k^{(o)} \triangleq X_k - X_k^{(r)}$ and *quantization error* $q_k^{(r)}$ as

$$\begin{aligned} Y_k &= X_k - \tilde{X}_k^{(r)} && (33)\\ &= X_k - X_k^{(r)} + q_k^{(r)} && (34)\\ &= Y_k^{(o)} + q_k^{(r)}. && (35) \end{aligned}$$

For a predictive block in a two-pass system, if the quantization parameters of its reference frame are $Q_1^{(r)}$ in the first pass and $Q_2^{(r)}$ in the second pass, then the corresponding DCT coefficients are

$$Y_k^{(1)} = Y_k^{(o)} + q_k^{(r)}\left(Q_1^{(r)}\right) \tag{36}$$

in the first pass and

$$Y_k^{(2)} = Y_k^{(o)} + q_k^{(r)}\left(Q_2^{(r)}\right) \tag{37}$$

in the second pass.

The motion estimation employed by MPEG encoders implies that it is valid to assume the open-loop prediction error and quantization error of the reference signal are uncorrelated as follows:

$$E\left[Y_k^{(o)} q_k^{(r)}\right] = E\left[Y_k^{(o)}\right] E\left[q_k^{(r)}\right]. \tag{38}$$

In addition, the quantization error is considered to be a zero-mean random variable

$$E\left[q_k^{(r)}\right] = 0. \tag{39}$$

We then find that the variances of $Y_k^{(1)}$ and $Y_k^{(2)}$ are related by

$$\sigma^2_{Y_k^{(1)}} - \sigma^2_{q_k^{(r)}\left(Q_2^{(r)}\right)} = \sigma^2_{Y_k^{(2)}} - \sigma^2_{q_k^{(r)}\left(Q_1^{(r)}\right)}. \tag{40}$$

TABLE VI
VNSMA-DCT FOR ZERO-MEAN GAUSSIAN, LAPLACIAN, AND UNIFORM RANDOM VARIABLES

	Gaussian	Laplacian	Uniform
ϕ	0.6366	0.5	0.75

MPEG uses uniform scalar quantizers and thus the variance of the quantization error is equal to the signal MSE

$$\sigma^2_{q_k^{(r)}\left(Q^{(r)}\right)} = D_k^{(r)}\left(Q^{(r)}\right) \tag{41}$$

where D_k is the MSE of the kth DCT coefficient. We then further rewrite (40) as

$$\sigma^2_{Y_k^{(2)}} = \sigma^2_{Y_k^{(1)}}\left(1 + \frac{D_k^{(r)}\left(Q_2^{(r)}\right) - D_k^{(r)}\left(Q_1^{(r)}\right)}{\sigma^2_{Y_k^{(1)}}}\right) \tag{42}$$

indicating that the variance of the DCT coefficient in the second pass can be estimated from the variance of the corresponding DCT coefficient in the first pass and the DCT MSEs $D_k^{(r)}(Q_1^{(r)})$ and $D_k^{(r)}(Q_2^{(r)})$.

Since finding the DCT MSE for the individual coefficients is a very complicated process, based on our experimental results, it is a reasonably good approximation to simplify (42) by replacing the exact DCT MSE by frame-average DCT MSE $\overline{D_k^{(r)}}(Q_1^{(r)})$ and $\overline{D_k^{(r)}}(Q_2^{(r)})$. Therefore, we use

$$\sigma^2_{Y_k^{(2)}} = \sigma^2_{Y_k^{(1)}}\left(1 + \frac{\overline{D_k^{(r)}}\left(Q_2^{(r)}\right) - \overline{D_k^{(r)}}\left(Q_1^{(r)}\right)}{\sigma^2_{Y_k^{(1)}}}\right) \tag{43}$$

in practice.

APPENDIX II
SADCT COMPENSATION FOR BETTER BIT RATE ESTIMATION

From Appendix I, we understand that the block statistics SADCT of the first pass have to be compensated in order to obtain more accurate bit-rate estimates in the second pass. We need to find out how to proceed from the variance relationship (43) to a meaningful relationship between the SADCT values.

For a random variable x, we define *variance-normalized square mean absolute* DCT (VNSMA-DCT) as

$$\phi \triangleq \frac{\left(\mathrm{E}\left[|x|\right]\right)^2}{\sigma_x^2}. \tag{44}$$

For Gaussian, Laplacian, and uniform random variables, ϕ can be calculated in closed-form, and the values are listed in Table VI. As experimental results show, in a real two-pass encoder, the value of ϕ will change in the second pass because of the changes in the statistics of the residual caused by the different reconstructed reference block that is used in the second pass when $Q_2^{(r)} \neq Q_1^{(r)}$. The implication of this is that different predictors lead to different distributions of the residual data. Two different variables $\phi_{1,k}$ and $\phi_{2,k}$ must be used in two different passes.

The ratio between the second- and first-pass DCT coefficients $\gamma_k \triangleq |C_k^{(2)}|/|C_k^{(1)}|$ can be estimated using

$$\gamma_k \doteq \frac{E\left[\left|C_k^{(2)}\right|\right]}{E\left[\left|C_k^{(1)}\right|\right]} \tag{45}$$

$$= \left[\frac{\phi_{2,k}}{\phi_{1,k}}\left(1 + \frac{D_k^{(r)}\left(Q_2^{(r)}\right) - D_k^{(r)}\left(Q_1^{(r)}\right)}{\sigma_k^2(Q_1)}\right)\right]^{\frac{1}{2}} \tag{46}$$

where $C_k^{(1)}$ and $C_k^{(2)}$ are the kth coefficients in the first and second passes, respectively. The remaining task is to estimate $\phi_{2,k}/\phi_{1,k}$. To further reduce complexity, the average values $\phi_i = 1/64\sum_{k=1}^{64} = \phi_{i,k}$, with $i = 1{,}2$, are used instead. The ratio ϕ_2/ϕ_1 is considered to be a function of $Q_2^{(r)}$, and $Q_1^{(r)}$ as

$$\frac{\phi_2}{\phi_1} = f_{\phi_2/\phi_1}\left(Q_2^{(r)}, Q_1^{(r)}\right). \tag{47}$$

There is no closed-form solution to this equation, and we model the function using training data. At the end of the first pass, a frame will be classified based on ϕ_1 and a precalculated f_{ϕ_2/ϕ_1} for the corresponding class is selected to form an estimate for ϕ_2. The second-pass SADCT can be computed using

$$\mathrm{SADCT}_2 = \sum_{k=1}^{64} \gamma_k \left|C_k^{(1)}\right|. \tag{48}$$

It is required to store all of the first-pass DCT coefficients in order to compute (48). In this work, we find the following simplification can be adopted without compromising much estimation accuracy as

$$\mathrm{SADCT}_2 = \hat{\gamma} \times \mathrm{SADCT}_1 \tag{49}$$

where

$$\hat{\gamma}\left(Q_2^{(r)}, Q_2^{(r)}\right) = \frac{1}{64}\sum_{k=1}^{64} \gamma_k. \tag{50}$$

ACKNOWLEDGMENT

The authors would like to thank J. Bankoski of On2 Technologies, Clifton Park, NY, for providing the movie clip *Stavro*.

REFERENCES

[1] *Generic Coding of Moving Pictures and Associated Audio Information: Video*, ISO/IEC 13818-2, MPEG Committee, May 1996.
[2] *Coding of Audio-Visual Objects—Part 2: Visual*, ISO/IEC JTC1, ISO/IEC 14496-2, 1999, (MPEG-4 visual version 1).
[3] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*. Englewood Cliffs, NJ: Prentice-Hall, 1984.
[4] J. Y. Huang and P. M. Schultheiss, "Block quantization of correlated Gaussian random variables," *IEEE Trans. Commun.*, vol. COM-11, no. 9, pp. 289–296, Sep. 1963.
[5] A. Segall, "Bit allocation and encoding for vector sources," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 2, pp. 162–169, Mar. 1976.
[6] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 9, pp. 1445–1453, Sep. 1988.

[7] A. Ortega, K. Ramchandran, and M. Vetterli, "Optimal buffer-constrained source quantization and fast approximations," in *Proc. ISCS*, May 1992, pp. 192–195.
[8] A. Ortega, "Optimal bit allocation under multiple rate constraints," in *Proc. DCC*, Mar. 1996, pp. 349–358.
[9] A. Ortega, K. Ramchandran, and M. Vetterli, "Optimal Trellis-based buffered compression and fast algorithms," *IEEE Trans. Image Process.*, vol. 3, no. 1, pp. 26–40, Jan. 1994.
[10] 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, no. 1, pp. 12–20, Feb. 1996.
[11] T. Chiang and Y.-Q. Zhang, "A new rate control scheme using quadratic rate distortion model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 2, pp. 246–250, Feb. 1997.
[12] L. J. Lin and A. Ortega, "Bit-rate control using piecewise approximated rate-distortion characteristics," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 6, pp. 446–459, Aug. 1998.
[13] Z. He and S. K. Mitra, "A linear source model and a unified rate control algorithm for DCT video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 11, pp. 970–982, Nov. 2002.
[14] H. M. Kim, "Adaptive rate control using nonlinear regression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 5, pp. 432–439, May 2003.
[15] 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, no. 2, pp. 287–298, Apr. 1997.
[16] J. J. Chen and H. M Hang, "Source model for transform video coder and its application—Part II: Variable frame rate coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 2, pp. 299–311, Apr. 1997.
[17] J. L. Mannos and D. J. Sakrison, "The effects of a visual fidelity criterion on the encoding of images," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 3, pp. 525–536, Jul. 1974.
[18] J. J. DePalma and E. M. Lowry, "Sine wave response of the visual system. II. Since wave and square wave contrast sensitivity," *J. Opt. Soc. Amer.*, vol. 52, pp. 328–335, Mar. 1962.
[19] N. B. Nill, "A visual model weighted cosine transform for image compression and quality assessment," *IEEE Trans. Commun.*, vol. COM-33, no. 6, pp. 551–557, Jun. 1985.
[20] A. Puri and R. Aravind, "Motion-compensated video coding with adaptive perceptual quantization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 1, no. 4, pp. 351–361, Dec. 1991.
[21] B. Tao, B. W. Dickinson, and H. A. Peterson, "Adaptive model-driven bit allocation for MPEG video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 1, pp. 147–157, Feb. 2000.
[22] K. Ramchandran, A. Ortega, and M. Vetterli, "Bit allocation for dependent quantization with applications to multiresolution and MPEG video coders," *IEEE Trans. Image Process.*, vol. 3, no. 9, pp. 533–545, Sep. 1994.
[23] K. Wang, "MPEG motion picture coding with long-term constraint on distortion variation," Ph.D. dissertation, Elect. Comput. Sci. Eng. Dept., Rensselaer Polytech. Inst., Troy, NY, May 2004.
[24] *Test Model 5*, ISO/IEC/JTC1/SC29/WG11, MPEG Committee, Apr. 1993.
[25] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ.Press, Mar. 2004.

Kai Wang (M'98) received the B.S. degree in electronic physics from Fudan University, Shanghai, China, in 1996, and the M.S. and Ph.D. degrees in electrical engineering from Rensselaer Polytechnic Institute, Troy, NY, in 2000 and 2004, respectively.

He was a summer intern with Sarnoff Corporation, Princeton, NJ, in 2000 and 2001. He is currently with Qualcomm Inc., San Diego, CA. His research interests include digital signal processing, image and video compression and transmission, and wireless communications.

John W. Woods (M'70–SM'83–F'88) received the B.S.E.E, M.S.E.E., E.E., and Ph.D. degrees from Massachusetts Institute of Technology, Cambridge, MA, in 1965, 1967, 1967, and 1970 respectively.

Since 1976, he has been with the Electrical Computer and Systems Engineering Department, Rensselaer Polytechnic Institute, Troy, NY, where he is currently a Professor and Director of the Center for Next Generation Video.

Dr. Woods is a member of Sigma Xi, Tau Beta Pi, Eta Kappa Nu, and the AAAS. He served on the editorial board of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY and was Chairman of the Seventh Workshop on Multidimensional Signal Processing in 1991. He served as Technical Program Co-Chairman for the first IEEE International Conference on Image Processing (ICIP), Austin, TX, November 1994. He was the co-recipient of the 1976 and 1987 Senior Paper Awards of what is now the IEEE Signal Processing (SP) Society, as well as the SP Society Meritorious Service Award in 1990, the SP Society Technical Achievement Award in 1994, and the IEEE Third Millennium Medal in 2000.