Title:

Operational Distortion-Quantization Curve Based Bit Allocation for Smooth Video Quality*

*This work was supported by NIH and NSF, under grant number NIH DHHS 1 R01 DC04340-01A2 and NSF Equipment Award: NSF EIA 9911095 respectively.

Authors:

Junqiang Lan, Wenjun Zeng, Xinhua Zhuang
{jl629@mizzou.edu; zengw@missouri.edu; zhuangx@missouri.edu}

Affiliation:

Computer Science Department
University of Missouri-Columbia
Columbia, MO, 65211

Manuscript correspondence:

Junqiang Lan
EBW 201, Computer Science Department,
University of Missouri Columbia, MO, 65211
Email: jl629@mizzou.edu
Tel: (01)573-882-0594
Fax: (01)573-882-8328

Abstract:

Quality fluctuation has a major negative effect on subjective video quality. A desirable single-pass frame-level constant-distortion bit allocation scheme is proposed in the paper for smooth video quality throughout the video sequence. The average distortion of all previous coded frames is taken as the target distortion for the current frame. According to the linear rate control algorithm [1][2], the distortion and rate are exponential and linear functions of the number of zero coefficients, respectively. Based on the distribution of the absolute DCT coefficients, an operational D-Q (Distortion-Quantization Stepsize) relationship curve can be established quickly. With the target distortion and operational D-Q curve, we then derive the close-form formulas for estimating the number of zero coefficients and the slope θ in the linear rate model, as well as the bit budget for the current frame. Experimental results show that the proposed constant-distortion (but variable rate) bit allocation scheme provides much smoother video quality on all test video sequences than the constant bit allocation scheme that has been used in many standard video coding reference software.

Key Words:

Bit Allocation, rate control, video coding, video streaming, smooth quality, operational distortion-quantization

Operational Distortion-Quantization Curve Based Bit Allocation for Smooth Video Quality

Junqiang Lan, Xinhua Zhuang, and Wenjun Zeng

I. INTRODUCTION

It is well known that bit rate control plays a key role in video coding. The goal is to achieve the best-perceived video quality under real-world constraints such as bandwidth, delay, and computation complexity. The channel bandwidth and delay constraints on bit rate control are both realized by introducing buffers as illustrated in Fig. 1, which shows how a typical video communication system works. The "channel" here could be either real communication channel or storage device. It can be seen that there are two video buffers in the path of data flow, one after the video encoder's output and the other before the decoder's input. These buffers are both circular FIFO (First-In-First-Out) queues with bits being written in and read out concurrently. In a constant bit rate transmission system, since the encoder instantaneously inserts compressed video frames to the buffer at a constant rate, while the decoder extracts the compressed frames at a similar constant rate, the sum of the encoder buffer delay and decoder buffer delay is a constant D [3].

$$B_e(t) + B_d(t + D) = RD \tag{1}$$

in which $B_e(t)$ and $B_d(t+D)$ are the buffer fullness of the encoder and decoder respectively, and R is the channel bandwidth. So normally the size of the encoder buffer is chosen as DR and the size of the decoder buffer is chosen as $DR + \delta$, where δ is a small positive number. For simplicity, the size of a buffer is often directly denoted as its corresponding delay time D. In interactive real-time video communications such as video conferencing or video phone, the buffer usually has very limited size due to the strict end-to-end delay requirement.

The encoder buffer is the bridge between the VBR (Variable Bit Rate) output of the video encoder and the channel transmission rate, which is considered as CBR (Constant Bit Rate) in most cases, but could also be considered as VBR under certain application scenarios (such as video transmission over wireless/Internet). At any moment, when the encoder generates more bits than the buffer could hold, a buffer *overflow* happens. The encoder then will either re-encode the current frame with coarser quantizers or simply drop it to avoid the overflow. This is why there is a feedback connection between the encoder buffer and the bit rate control module in Fig. 1. At any moment when there are no bits available in the encoder buffer, we call it a buffer *underflow*. Buffer underflow wastes some of the available bandwidth. Because of the relationship between the encoder buffer fullness and the decoder buffer fullness in Eq. (1), encoder overflow implies a delayed underflow at decoder side and vice versa. In order to meet the channel bandwidth and delay constraints, the bit rate control at encoder side needs to ensure that there is neither underflow nor overflow at any time.

The actual bit rate control algorithm normally includes two parts: bit allocation and quantization control. In bit allocation, the target bit budget is allocated among different coding units such as GOPs (Group of Pictures), frames, slices, video objects or macroblocks. In quantization control, optimal quantization parameter Q is determined to achieve the allocated bit budget. Both of these two parts have been extensively studied under the framework of rate-distortion theory.

For the determination of the Q, the R-Q relationship has to be known so that the appropriate Q could be chosen for the specific bit rate R. The R-Q curve generally depends on both the current (residual) video frame and the encoder algorithm; and it is highly non-linear. The curve can be plotted either by model-based method or by operational method. In operational solutions, we enumerate all the possible Qs and get their corresponding Rs by actually encoding the frame. It is obvious that operational R-Q curve is most accurate, but its prohibitively high computational complexity makes it rarely adopted in practical, especially real time video coding applications. In model based solutions, however, R-Q relation is approximated with some close-form mathematical functions such as power[4], logarithmic [5][6], polynomial[7][8][9], spline [10], exponential[11] or even more complicated [12], with several parameters to be determined. The estimation of R-Q relationship is then simplified to a parameter estimation problem with the statistical knowledge of previous encoding output and the current video frame. Recently, He and Mitra [1][2] proposed the ρ domain rate control method, in which ρ is the percentage of zeros among the quantized DCT coefficients. In their work, instead of explicitly figuring out the R-Q curve, they consider both R and Q as the function of ρ. With a determined one-to-one mapping between Q and ρ given the frame's DCT coefficient distribution and a linear relationship between R and ρ, the bit rate control model is simplified to the estimation of the slope in the linear R-ρ relation. But due to the non-stationary nature of the video sequences, model inaccuracy exists in all the model based solutions.

For the bit allocation part, a lot of video coding applications pick the "average" bit allocation scheme, in which each frame receives the same allocated bit budget, i.e. the target bit rate divided by the target frame rate. This constant bit allocation is adopted sometimes because of the small buffer size required by real time interactive applications, and sometimes just because of its simplicity. The constant bit allocation is widely deployed in standard reference video codec software such as TM5 for MPEG2 [13], MPEG4 Annex L [14][15] and TMN8 for H.263+[16][17]. All these constant bit allocation are based on the assumption that video content is stationary across scenes and frames. This assumption often does not hold in natural video sequences, and thus the constant bit allocation solutions need to be improved to achieve smooth video quality.

In video streaming applications, with a reasonable larger buffer size, it is possible to alleviate the quality fluctuation problem using content adaptive bit allocation scheme. In [18] - [22], several non-uniform bit allocation schemes are proposed for coding frames within a GOP (fixed size group of pictures). However, the bit allocation among GOPs remains constant and inconsistent video quality among GOPs still exists. Xie and Zeng [23] recently proposed to use MAD (Mean Absolute Difference) as the video content complexity measurement and allocate bit budget based on the MAD value of current frame. They start from constant Q encoding for the entire sequence, which is considered as a fairly good model of constant video quality. Based on the observation that R-MAD curve is very close to a quadratic function under constant Q, they approximate the relation between the output bit rate R and the MAD as a quadratic function. The bit budget is then allocated proportional to the square root of the ratio between current MAD and the average MAD of all previous encoded frames. Their algorithm has extremely low complexity and has better performance than the constant bit allocation scheme. But their solution starts from constant Q and the objective of the optimization is not explicitly defined. Another quality smoothing algorithm, recently proposed by Chen and He [24] , directly targets on constant distortion. The average distortion of previous 50 encoded frames is taken as the target distortion of the current fame. Based on the *D-Q-R* relation model first proposed in TMN8 [16][25], an iterative algorithm is formulated to calculate the target bit rate of the current frame that is expected to yield the desired distortion. Their solution has an explicit objective, but the *D-Q* model ($D = \frac{Q^2}{12}$) they adopted is based on the assumption that DCT coefficients are uniformly distributed over each interval of length Q. But it is well known that DCT coefficients are more likely to be Laplacian distributed [26]. The inaccurate *D-Q* model may cause inaccurate bit budget allocation and in turn may lead to a distortion value away from the target distortion. That is why they introduced an iterative algorithm that intends to make the actual distortion converge to the desired distortion.

In this paper, a novel single-pass frame-level constant-distortion bit allocation scheme is proposed. Similar to [24], we explicitly target on constant distortion. But our target distortion of the current frame is set as the average distortion of all previous encoded frames, instead of the windowed (50 previous frame) average. We then adopt the linear rate model proposed in [1][2] to estimate the target bit budget for the current frame in our solution. We propose to first use the operational D-Q curve to quickly localize the estimate in a very small Q (or ρ) interval. Then the linear rate model is applied to the identified small region so that the parameter estimation will be more reliable/accurate. Close-form formulas for estimating the number of zero coefficients and the slope θ in the linear rate model, as well as the bit budget for the current frame are derived. Once the bit budget for the current frame is determined, any macroblock layer rate control algorithms, e.g. the one used in TMN8 can be used to determine the Q for each MB. Experiment results show that our algorithm significant alleviate the quality fluctuation and has better encoded video quality.

The paper is organized as follows. In section II, we will briefly review two related works, the standard H.263+ constant rate control algorithm TMN8 [16], which is known as one of the best quantization determination schemes, and the linear rate control algorithm, which was proposed by He and Mitra [1][2]. Both algorithms were not designed for the frame-level bit allocation.

Section III explains our proposed operational *D-Q* model. The proposed frame-level constant-distortion bit allocation algorithm is discussed in details in Section IV. Performance evaluation is given in Section V. In Section VI, we make our conclusion.

II. RELATED WORKS

TMN8 [16] is a model-based rate-distortion optimized bit rate control algorithm, in which the target frame bit rate is constant. In TMN8, both the rate *R* and distortion *D* are taken as the function of quantization stepsize *Q*.

$$D = \frac{1}{N}\sum_{i=1}^{N}\frac{Q_i^2}{12}$$
$$B_i = A(K\frac{\sigma_i^2}{Q_i^2} + C) \tag{2}$$

where *D* is the distortion; Q_i is the quantization step size; *N* is the number of macroblocks (MBs) in a frame; A is the number of pixels in a macroblock; σ_i^2 is the macroblock's standard deviation and *C* is the overhead (such as header, motion vectors and etc) rate. TMN8 selects optimized quantization step-sizes for all the macroblocks to minimize the overall distortion while meeting the constant target frame bit rate. The problem is formulated as a constrained optimization problem and the solution to the Lagrange multiplier optimization, i.e. the optimized *Q**, is expressed as:

$$Q_i^* = \sqrt{\frac{AK}{(\beta_i - ANC)}\frac{\sigma_i}{\alpha_i}\sum_{k=1}^{N}\alpha_i\sigma_i} \tag{3}$$

in which β_i is the target bit rate for all the remaining uncoded macroblocks; *N* is the number of remaining uncoded macroblocks; α_i is the distortion weight of each macroblock. For details about the model and the rate control algorithm, the readers are referred to [25].

In the linear (ρ domain) rate control algorithm, the bit rate is observed to have a linear relationship with ρ , and the distortion has shown to be an exponential relationship with ρ [2]:

$$R = \theta \cdot (1-\rho)$$
$$D = \sigma^2 \cdot e^{-\alpha(1-\rho)} \tag{4}$$

where σ^2 is the variance of a video frame; θ and α are constants for a frame. When the distribution of the DCT coefficients is known, there exists a one-to-one mapping between ρ and the quantization stepsize *Q*:

$$\rho = \frac{1}{N}[\sum_{|x|<Q} D_0(x) + \sum_{|x|<1.25Q} D_1(x)] \tag{5}$$

where D_o and D_I are the coefficient distribution of INTRA and INTER coded blocks and N_t is the total number of coefficients. This one-to-one relationship enables the ρ domain algorithm to properly select Q in video coding. If we denote the number of zero coefficients as $\hat{N}$, Eq. (4) becomes

$$\begin{aligned} R &= \theta \cdot (N_t - \hat{N}) \\ D &= \sigma^2 \cdot e^{-\beta(N_t - \hat{N})} \end{aligned} \tag{6}$$

where for notational simplicity we use θ for θ / N.

The ρ domain rate-distortion optimization has been successfully applied to the optimal bit allocation within a video frame, including bit allocation among video objects or different groups of macroblocks within a frame [2].

III. OPERATIONAL D-Q CURVE

In a video coding system, the only lossy part is quantization. Because DCT is an orthonormal transformation, the MSE (Mean Square Error) distortion in spatial domain is the same as the MSE distortion in frequency domain. The actual distortion corresponding to a specific Q can be calculated as:

$$D(Q) = \sum_{u=0}^{W \times H \times 3/2-1} [x(u) - r_Q(x(u))]^2 \tag{7}$$

in which x is the original DCT coefficient value; $r_Q(x)$ is the reconstruction value of x defined by the quantization scheme; W and H is the image width and height respectively; the factor 3/2 is due to the fact that both luminance and chrominance need to be counted. To get one reconstruction value, a quantization and dequantization operation is needed. So totally the distortion for one specific Q needs *WxHx3/2* times quantization /dequantization operations for each frame.

In this paper, we propose to use coefficient distribution to simplify the distortion calculation so that all the quantization/dequantization operations for the same coefficient value only need to be executed once. Denote the histogram of the absolute DCT coefficients by H. The distortion caused by quantization Q can be calculated as:

$$D(Q) = \sum_{x=0}^{M} \int_{x-0.5}^{x+0.5} (y - r_Q(x))^2 \, p(y) dy \tag{8}$$

where x is the rounded integer absolute DCT coefficient value; $r_Q(x)$ is still the reconstruction value of x; M is the number of entries of the histogram; $p(y)$ is the underlying probability distribution of the absolute DCT coefficients and $\int_{x-0.5}^{x+0.5} p(y) dy = \frac{H(x)}{N}$;

N is the total number of DCT coefficients, and $N = \sum_{x=0}^{M} H(x)$. The DCT coefficients are normally modeled by Laplacian distribution [26]. Here we approximate *p(y)* as uniformly distributed inside each relatively small interval [*x-0.5*, *x+0.5*]. Thus the distortion in Eq. (8) can be approximated as:

$$D(Q) = \frac{1}{N}\sum_{x=0}^{M}(x - r_Q(x))^2 H(x) + \frac{1}{12} \tag{9}$$

As can be seen, the entire *D-Q* curve can be determined by Eq. (9) once the histogram H is counted and the reconstruction function $r_Q(x)$ is defined. In practice we observed that more than 90% of the DCT coefficient is less than 128, so we set M=128 and the average distortion contributed by all the values larger than 128 is approximated as $\frac{Q^2}{12}$. This is because the distribution in the [128, ∞) interval usually looks flat and can be reasonably approximated as uniform distribution. Then Eq. (9) becomes:

$$D(Q) = \frac{1}{N}\left(\sum_{x=1}^{128}(x - r_Q(x))^2 H(x) + L\frac{Q^2}{12}\right) + \frac{1}{12} \tag{10}$$

where L is the number of coefficients larger than 128. Since the distortion of value 0 is also zero, so *x=0* is excluded from the summation. Please note that Eq. (10) differs from the accurate *D(Q)* defined in Eq. (7) in that we use the following three approximations:

1. Rounding all the DCT coefficients to integer values first. This is in fact a quantization with step size 1.
2. Assuming that the distribution in each [x-0.5, x+0.5] interval is uniform; the distortion caused by the rounding and this uniform distribution assumption is finally approximated as $\frac{1}{12}$.
3. Approximate the average distortion coming from all the values larger than 128 as $\frac{Q^2}{12}$.

From the experimental data in Fig. 2, we can see that the *D-Q* curve determined by Eq. (10) is very close to the *D-Q* curve generated by actually encoding the frame using Q (or Eq. (7)). However, Eq. (10) involves only 128 quantization/ dequantization operations for any Q, which is $\frac{128}{W \times H \times 3/2}$ of the complexity of Eq. (7). For example, for a QCIF video the complexity of Eq. (10) is only 0.34% of Eq (7).

IV. FRAME-LEVEL CONSTANT-DISTORTION BIT ALLOCATION (CDBA)

The goal of the proposed algorithm is to minimize the distortion variation throughout the whole video sequence. The proposed algorithm works as follow. Prior to the current frame being encoded, the average distortion of all the previous coded frames is calculated and taken as the target distortion. Then the number of zero DCT coefficients, $\hat{N}$, which corresponds to the target distortion, is estimated. Next the slope θ of the linear rate control algorithm is estimated. Based on the estimated θ and the number of zeros, $\hat{N}$, a bit budget is allocated. Finally, the allocated bit budget is shaped by the buffer constraints to guarantee there is no overflow and underflow at any time. Thus, the proposed algorithm provides constant- or nearly constant-distortion bit allocation.

The average distortion of (K-1) previous encoded frames is calculated as:

$$\hat{D}_K = \frac{1}{K-1}\sum_{i=1}^{K-1} D_i \tag{11}$$

where D_i is the distortion measured at frame *i.* The distortion metric adopted is MSE (Mean Square Error), which directly reflects the objective video quality in terms of PSNR (Peak Signal to Noise Ratio). Ideally the target distortion should be the average distortion in the same scene since the pictures in the same scene tend to have similar rate-distortion characteristics. So a fixed windowed average distortion as adopted in [24] may include some scene change frames and the possible result is that the target distortion itself is fluctuating and may cause man-made quality fluctuation. So here the average of all previous encoded frame is introduced to stabilize the target distortion and thus would reduce distortion fluctuation in the long run.

A. *Estimate of Number of Zeros*

There are only two unknown parameters (σ^2 and β) in the *D-N* (distortion-number of zeros) model in Eq. (6), so any two known (D, N) pairs would solve the equation and determine the D-N model. Then number of zeros, $\hat{N}$ corresponding to the average distortion $\hat{D}_K$ can be calculated immediately according to the determined model. But different *D-N* pairs will result in different parameter values. In order to avoid the model inaccuracy problem, we choose the two known *(D, N)* pairs which are nearest to ($\hat{D}_K$, $\hat{N}$) as the model reference points. Since *D(Q)* is a monotonically increasing function, without loss of generality we may assume that for some Q^*, $D(Q^*-1) < \hat{D}_K \le D(Q^*)$. If the bit rate control algorithm works only on frame level, i.e. all the macroblocks in the frame share the same *Q*, then $D(Q^*)$ or $D(Q^*-1)$ would be the best approximation of $\hat{D}_K$. This best approximation, however, may still deviate from $\hat{D}_K$ significantly, especially when Q^* is relatively small where the *D-Q* slope is large (see. Fig. 2). Accurate rate control algorithms such as TMN8 often work on macroblock layer so that each macroblock could have different *Qs*. In such cases, to minimize the quality fluctuation across frames, we need to find out the average fractional *Q'* or equivalently, the number of zeros corresponding to the target distortion of the frame. Denote the number of zeros for a specific *Q* as *N(Q)*. From Eq. (6), we have:

$$\begin{aligned} D(Q^*) &= \sigma^2 e^{-\beta(N-N(Q^*))} \\ D(Q^*-1) &= \sigma^2 e^{-\beta(N-N(Q^*-1))} \\ \hat{D}_K &= \sigma^2 e^{-\beta(N-\hat{N})} \end{aligned} \tag{12}$$

From the three equations in Eq. (12), the number of zeros $\hat{N}$, can be solved in close-form without explicitly knowing σ^2 and β as:

$$\hat{N} = \frac{N(Q^*-1)\ln(\frac{D(Q^*)}{\hat{D}_K}) + N(Q^*)\ln(\frac{\hat{D}_K}{D(Q^*-1)})}{\ln(\frac{D(Q^*)}{D(Q^*-1)})} \tag{13}$$

It can be seen that $\hat{N}$ is the linear combination of *N(Q*)* and *N(Q*-1)*. The weights are determined by three log-ratio functions involving the target distortion $\hat{D}_K$ and two bounding reference values in the *D-Q* function.

Because *D(Q)* is a monotonically increasing function, we can adopt the binary search algorithm [27] to find out *Q** from all the possible *Q*s, using our proposed simplified operational *D-Q* determination. The number of search points to find out *Q** in all the possible *Q*'s is $log(N_q)$, in which N_q is the number of possible *Qs*. For example, in H.263, N_q is 31, which means we only need maximal five *D(Q)* values in order to locate *Q**.

B. Estimate of θ

In MPEG video coders, the quantized DCT coefficients are zigzag organized/scanned, run length coded, and then entropy coded using variable length code. In essence, the algorithm encodes nonzero DCT coefficients by encoding (run, level) pairs, in which "run" is the number of zeros between the current nonzero coefficient and the previous nonzero coefficient and "level" is the current coefficient value. Thus, the (run, level) pair describes the position and the value of each nonzero DCT coefficient. At a different quantization level, the number $\hat{N}$ of zero coefficients may change but the cost of encoding nonzero DCT coefficients, or (run, level) pairs, i.e., θ, remains constant – as observed in [1][2] and summarized as the linear rate control algorithm. Qualitatively we may explain the linear relationship between the rate and the number of zeros as follows. At a coarser quantization scale, the quantized value (i.e., "level") decreases while the number of zeros ("run") increases. The overall effect of the increased "run" and decreased "level" "magically" results in an unchanged average cost for encoding a nonzero DCT coefficient.

To estimate the slope θ , we select a typical quantization stepsize Q_o, count the number N_o of zeros, and calculate the (run, level) pairs and their corresponding Huffman code lengths by look-up-table. At the end, we obtain the bit stream length R_o, and the slope θ is estimated as

$$\theta = \frac{R_o}{N - N_o} \tag{14}$$

The implementation of this part can be done together with the distribution collection. Since the histogram or distribution of the DCT coefficient is needed in order to get the target zero number. During the process of distribution collection, the coefficients in each 8x8 block is visited in the zig-zag scan order so that it is easy to know the (run, level) pair and find out the final bit rate with a specified quantization parameter Q_o. Here again there exists multiple choices of Q_o. According to our observation, the linear rate slope θ is relatively robust and not sensitive to different choice of Q_o compared with (σ^2 and β) in (5). So in our experiment we set Q_o as 10.

C. Bit Allocation

After both the slope θ and the number $\hat{N}$ of zeros are estimated, the bit budget $\widehat{R}$ to meet the target distortion $\hat{D}_K$ can be determined as

$$\hat{R} = \theta(N - \hat{N}) \tag{15}$$

Since the estimated bit budget $\widehat{R}$ only counts the bits for encoding DCT coefficients, headers and motion vectors will add more bits in the allocated bit budget as follows.

$$\hat{R} = \theta(N - \hat{N}) + R_{MV} + R_H \tag{16}$$

where R_{MV} and R_H are the bit budgets allocated for motion vectors and headers. R_{MV} cannot be determined accurately ahead of time because of the existence of "skip" mode, which signals the whole macroblock as one bit when the whole block has zero motion and all zero coefficients. Since the number of skipped blocks can be determined only after quantization stepsize is determined, R_{MV} can only be estimation before actual quantization.

To avoid buffer underflow and overflow, the allocated bit budget $\widehat{R}$ by Eq. (16) needs to be further shaped using buffer constraints, and the final bit budget allocation scheme is given by:

$$R_t = \begin{cases} \frac{C}{F} + T_o \cdot B_{\max} - B & if(\hat{R} + B - \frac{C}{F}) > T_o \cdot B_{\max} \\ \frac{C}{F} + T_u \cdot B_{\max} - B) & if(\hat{R} + B - \frac{C}{F}) < T_u \cdot B_{\max} \\ \hat{R} & otherwise \end{cases} \quad (17)$$

where F is the target frame rate; C is the target bit rate; B_{max} is the buffer size; B is the number of bits currently in the buffer; T_o and T_u are the overflow and underflow threshold respectively, which are set to 0.9 and 0.1 in the simulation. Eq. (16) and (17) together determine the final bit rate allocated for the current frame.

V. Performance Evaluation

We applied the proposed CDBA (Constant Distortion Bit Allocation) algorithm to H.263-based video streaming applications. In the implementation, the INTRA and INTER blocks produce separate histograms and contribute the distortions based on different reconstruction functions because the dead zone and quantization formula for INTRA and INTER macroblocks are different.

We compare our CDBA scheme with the following two bit allocation schemes: the MAD based bit allocation proposed by Xie and Zeng in [23] and the constant bit allocation scheme. The constant bit allocation is defined as:

$$\hat{R} = \frac{C}{F} - \Delta$$
$$\Delta = \begin{cases} B/F & if(B > 0) \\ \frac{(R_{K-1} - \hat{R}_{K-1})}{K-1} & Otherwise \end{cases} \quad (18)$$

where Δ is the accumulated discrepancy between the allocated bits number $\hat{R}_{K-1}$ and the actual bits number R_{K-1} for all already encoded frames. The reason why we do not adopt the bit allocation scheme in TMN8 is that in TMN8, the buffer size is only as small as *C/F*, the average bit budget of one frame. Whenever the output bit rate is bigger than this virtual buffer size, frames are skipped until the bits in the buffer are smaller than the buffer size. This frame skipping and strict constant bit allocation is only suitable in interactive low delay video communication applications. Since our target application is video streaming and no frame skipping is preferred, to be fair, we assume the same relative larger buffer size and the constant bit allocation algorithm is modified to Eq. (18).

The MAD based allocation algorithm is defined as:

$$\hat{R} = \frac{C}{F}\sqrt{\frac{MAD_K}{\overline{MAD}_{K-1}}} - \Delta \quad (19)$$

in which Δ is the same as in Eq. (18); MAD_K is the Mean Absolute Difference of current frame; $\overline{MAD}_{K-1}$ is the average MAD of previous K-1 frames. The buffer constraints for both algorithms are the same as defined in Eq. (17).

In our proposed algorithm, the bit allocation of the first 10 frames uses the constant allocation to achieve a stable initial average distortion. TMN8 macroblock layer rate control algorithm is adopted to determine the Q for each MB for all three bit allocation schemes. The buffer size is one second's length i.e. $B_{max} = C$, and the thresholds for underflow and overflow are 0.1 and 0.9, respectively.

Extensive simulations were performed on many standard sequences. Here we present a subset of representative results. The selected sequences are (all QCIF) News, Mother and Daughter, Foreman and a combined sequence of M&D, Akiyo and Car phone. The target bit rate and frame rate are all set to 48Kbps@10fps.

The output bit rate may have control errors depending on how many bits are left in the buffer at the end of encoding. The control error is bounded by:

$$\frac{0.1 \cdot C}{L} = \frac{T_u \cdot B_{\max}}{L} \leq E \leq \frac{T_o \cdot B_{\max}}{L} = \frac{0.9 \cdot C}{L} \tag{20}$$

in which L is the whole video duration in seconds. It is clear that the bit rate error becomes negligible when the sequence has much longer duration than the buffer size (in seconds). But for a three-second video sequence as in our simulation, the control error could be relatively large. In such cases, we re-adjust the target bit rate so that the actual output bit rates of all the algorithms remain the same.

From Table 1, it can be observed that at the same actual output bit rate, in terms of the output video quality measured by the average PSNR value, for small motion sequences such as "Mother & Daughter" or news, CDBA has comparable PSNR performance as the other two. For large motion (Foreman) and scene change video sequence (combined), CDBA always has comparable average video quality as the better one of the other two. But in terms of video quality fluctuation measured by the standard deviation of the quality of the video sequence, our proposed CDBA algorithm has the smallest standard deviation, which means least quality fluctuation.

Fig. 3 further illustrates the much smoother video quality of the proposed CDBA. In the "M&D" sequence, there is only small motion. It can be seen in the middle of the sequence, our proposed CDBA achieves almost constant distortion (PSNR) performance. In both "News" and the combined video sequence, there are several explicit scene change frames which correspond to the sharp PSNR drops of the constant allocation scheme (dotted lines). Both the MAD and the CDBA algorithm hold well at those scene change frames. From Fig. 3(b) and (c), we can see that CDBA has the smallest PSNR drops among the three algorithms. In Fig. 4, the bit rate consumed by each frame is plotted for the "combined sequence", we can see that at the scene change frames, both CDBA and MAD based approaches allocated much more bits to keep the smooth video quality. CDBA allocated a little more bits than MAD based approach to keep the constant distortion, which also corresponds to the least PSNR drop in Fig. 3 (c)

In fact, both CDBA and MAD based allocations appear to have much better subjective video quality since viewers usually tend to rate the video quality along the worst frame. Subjective video quality is compared in Fig. 5 for the three algorithms. The frames compared are two scene change frames in the combined sequence. The advantage of the proposed CDBA scheme is obvious.

On the other hand, the proposed CDBA scheme has the highest computational complexity among the three algorithms, yet the increased complexity is still affordable in a video coding system. There is no memory/computational overhead involved in constant bit allocation. In MAD based method, since the MAD of current frame needs to be known ahead of bit allocation, so it requires that motion estimation is first applied on the whole video frame first and the residual image is then stored. In CDBA based method, since the histogram of DCT coefficients is needed to derive the bit budget, both motion estimation and DCT transform on the whole frame are applied first and the DCT coefficients are stored instead of the spatial residual image itself. So in term of memory consumption, our approach has the same memory requirement with the MAD method. The difference is that MAD stores the spatial domain image while CDBA stores DCT domain image although both spatial and DCT domain residual image needs 16 bits/pixel. The computation part of CDBA algorithm has higher complexity than MAD based algorithm because we need to count the histogram of DCT coefficients first and then figure out the operational *D-Q* curve. But the worst case computational cost for the operational *D-Q* curve only needs ten 8x8 blocks quantization/dequantization. The reason is that using binary search (section IV.A) we need at most five *D-Q* pairs to find Q^* and each *D-Q* pair involves 128 (two 8x8 blocks) values' quantization/dequantization. With the fact that an MB (Macro Block) normally has six 8x8 blocks, the computational cost for the *D-Q* curve only needs less than two MB's quantization and dequantization, which is definitely affordable compared with the huge computation involved in the video coding itself.

VI. CONCLUSION

In this paper, we have proposed a novel frame level single pass bit allocation algorithm for smooth video quality in video streaming applications. Experimental results show that it greatly alleviates the quality fluctuation problem observed in constant bit allocation. There are two major contributions of the work. First, we propose to use a simplified operational Distortion-Quantization approach with affordable complexity to avoid the model inaccuracy problem that exists in all the model based solutions. The second contribution is that we propose a practical way of estimating the rate slope in ρ domain linear rate control algorithm. This slope estimation could be used in ρ domain rate control as the initial slope value. The algorithm can be extended to all MPEG like video coders to alleviate the quality fluctuation problem.

REFERENCES

[1] Z. He and S.K. Mitra, “A linear source model and a unified rate control algorithm for DCT video coding”, IEEE Transactions on Circuits and System for Video Technology. vol. 12, pp. 970-982, Nov. 2002.

[2] Z. He and S.K. Mitra, “Optimum bit allocation and accurate rate control for video coding via ρ domain source modeling,” IEEE Transactions on Circuits and System for Video Technology. Volume: 12, pp.840-849, Oct. 2002

[3] P. Assunção and M. Ghanbari, “Buffer Analysis and Control in CBR Video Transcoding,” IEEE Transactions on Circuits and System for Video Technology, Vol. 10, No. 1, pp. 83-92, Feb. 2000.

[4] W. Ding and B. Liu, “Rate control of MPEG video coding and recording by rate-quantization modeling,” IEEE Transactions on Circuits and System for Video Technology, vol. 6, pp. 12-20, Feb. 1996.

[5] H.-M. Hang and J.-J. Chen, “Source model for transform video coder and its application – part I: fundamental theory,” IEEE Transactions on Circuits and System for Video Technology, vol. 7, pp. 287-298, Apr. 1997.

[6] E. D. Frimout, J. Biemond and R. L. Lagendik, “Forward rate control for MPEG recording,” in Proc. SPIE Visual Communication Image Processing, Cambridge, MA, Nov. 1993, pp. 184-194.

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

[8] J. I. Ronda, M. Eckert, F. Jaureguizar and N. Garcia, “Rate control and bit allocation for MPEG-4,” IEEE Transactions on Circuits and System for Video Technology, vol. 9, no. 8, pp. 1243-1258, Dec. 1999.

[9] L.-J. Lin and A. Ortega, “Bit-rate control using piecewise approximated rate-distortion characteristics,” IEEE Transactions on Circuits and System for Video Technology, vol. 8, no. 4, pp. 446-459, Aug. 1998.

[10] L.-J. Lin, A. Ortega and C.-C. J. Kuo, “Rate control using spline-interpolated R-D characteristics,” SPIE Visual Communication Image Processing, Cambridge Visual Communication Image Processing, Cambridge, Orlando, FL, Mar. 1996, pp. 111-122.

[11] J. Bai, Q. Liao, X. Lin, and X. Zhuang, “Rate –distortion model based rate control for real-time VBR video coding and low-delay communications”, Signal Processing: Image Communication, vol. 17, no. 2, Feb. 2002, pp. 187-199.

[12] Y. Yang and S. S. Hemami, “Rate control for VBR over ATM: simplification and implementation," IEEE Transactions on Circuits and System for Video Technology, vol. 11, no.9, pp. 1045-1058, Sept.2001.

[13] “Test Model 5”, Draft, ISO-IEC/JTC1/SC29/WG11/N0400, April 1993.

[14] Coding of Moving Pictures and Associated Audio, MPEG-4 Committee Draft, ISO/IEC 14496-2, Oct. 1998.

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

[16] ITU-T SG16/Q15, Video Coding Expert Group, "Video Codec Test Model TMN8," Portland, June 1997.

[17] G. Cote, B. Erol, M. Gallant, and F. Kossentini, "H.263+: Video coding at low bit rates," IEEE Transactions on Circuits and System for Video Technology, vol. 8, pp.849–866, Nov. 1998.

[18] N. Mohawnian, R. Rajagopalan and C. A. Gonzales, "Single-pass constant- and variable-bit-rate MPEG-2 video compression," IBM J. Res. Develop. 43, No. 4, pp. 489-509, July 1999.

[19] P. H. Westerink, R. Rajagopalan, and C. A. Gonzales, "Two-pass MPEG-2 variable-bit-rate encoding," IBM J. Res. Develop. 43, No. 4, pp. 471-488, July 1999.

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

[21] K. Ramchandran, A. Ortega and M. Vetterli, "Bit allocation for dependent quantization with applications to multiresolution and MPEG video coders," IEEE Transactions on Image Processing, vol. 3,no. 5, pp. 533-544, Sep. 1994.

[22] A. Ortega, K. Ramchandran and M. Vetterli, "Optimal trellis-based buffered compression and fast approximations," IEEE Transactions on Image Processing, vol. 3, no. 1, pp. 26-40, Jan. 1994.

[23] B. Xie and W. Zeng, "Sequence-based rate control for constant quality video," in Proc. 2002 International Conference on Image Processing, Rochester, New York, Sept. 2002.

[24] T. Chen and Z. He, "Signal-pass distortion-smoothing encoding for low bit-rate video streaming applications," International Conference on Multimedia and Expo, Baltimore, USA, Jul. 2003.

[25] J. Ribas-Corbera and S. Lei, "Rate control in DCT video coding for low-delay video communications," IEEE Transactions on Circuits and System for Video Technology, vol. 9, Feb. 1999, pp. 172–185.

[26] Smoot S. R. and Rowe L. "A Study of DCT coefficient distributions," Proceedings of the SPIE Symposium on Electronic Imaging, volume 2657, San Jose, CA, Jan. 1996.

[27] T. H. Cormen, C. E. Leiserson, R. L. Rivest and C. Stein, "Introduction to Algorithms", MIT Press, 2nd Edition, 2001.

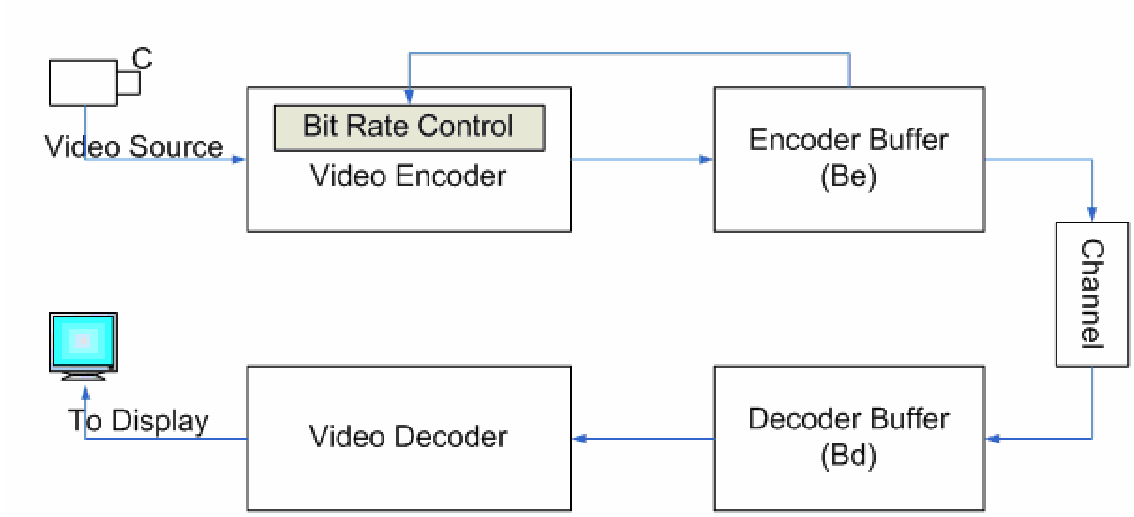


Figure1. Diagram of a typical video transmission system

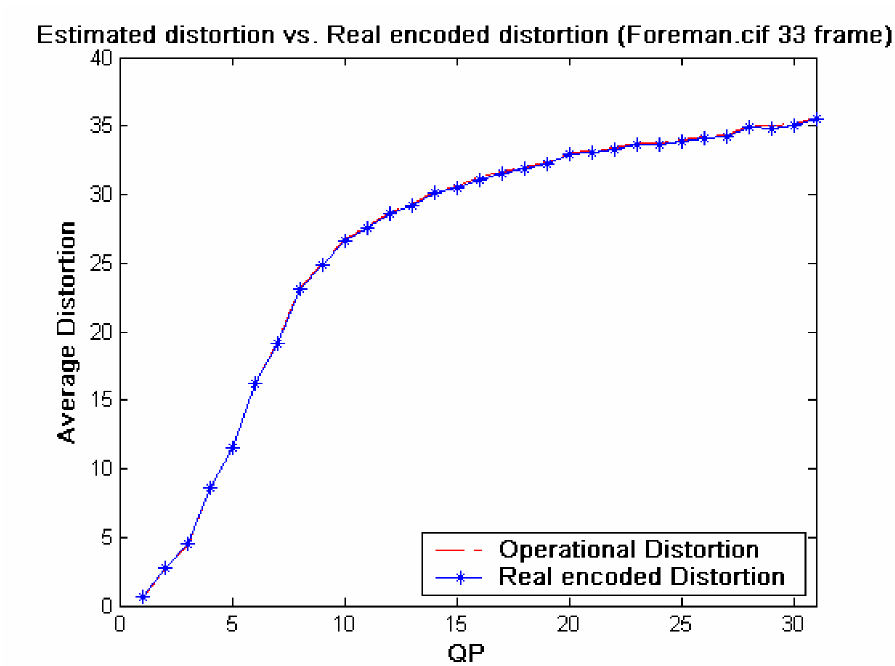
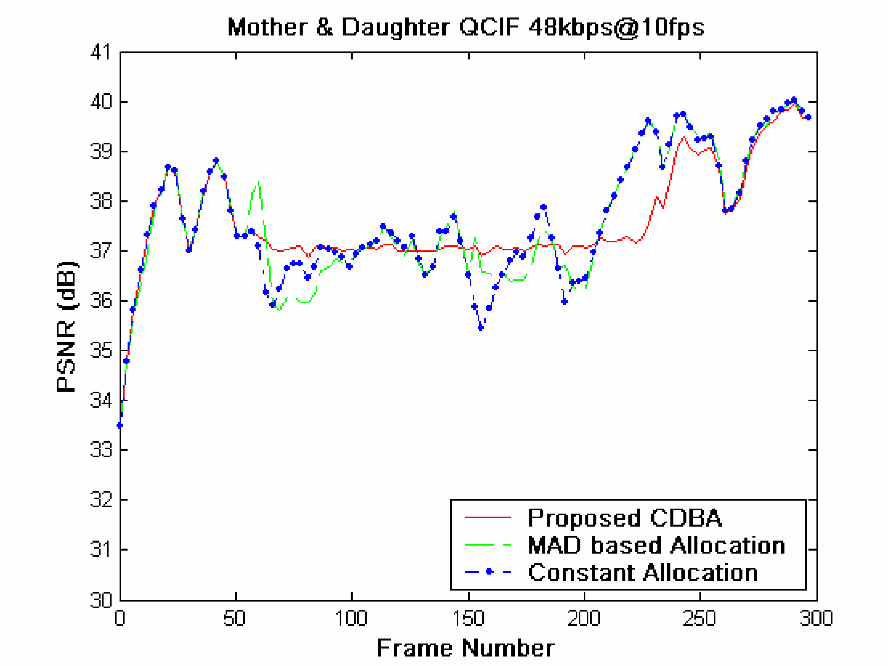
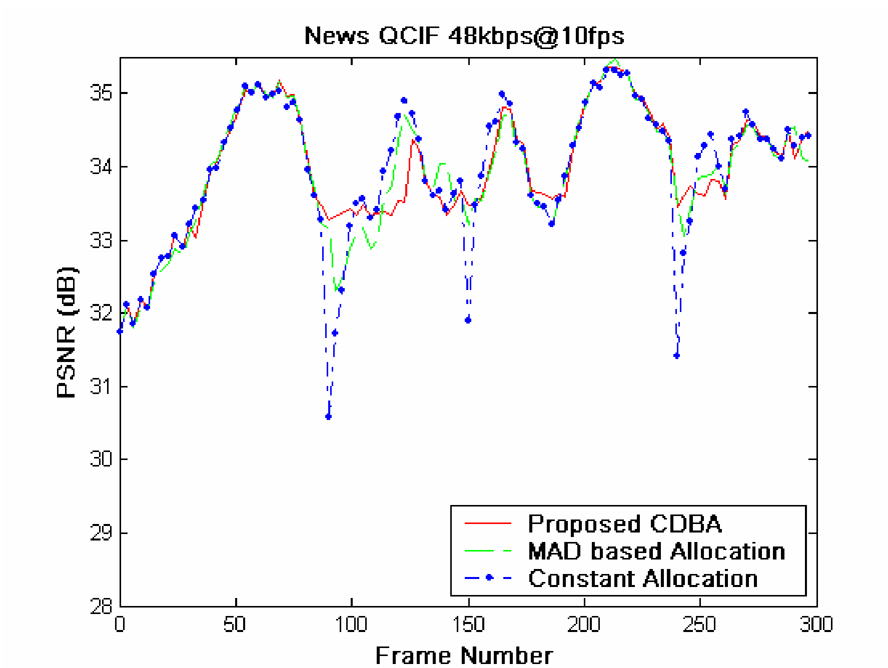


Figure 2. Estimated distortion vs. real distortion



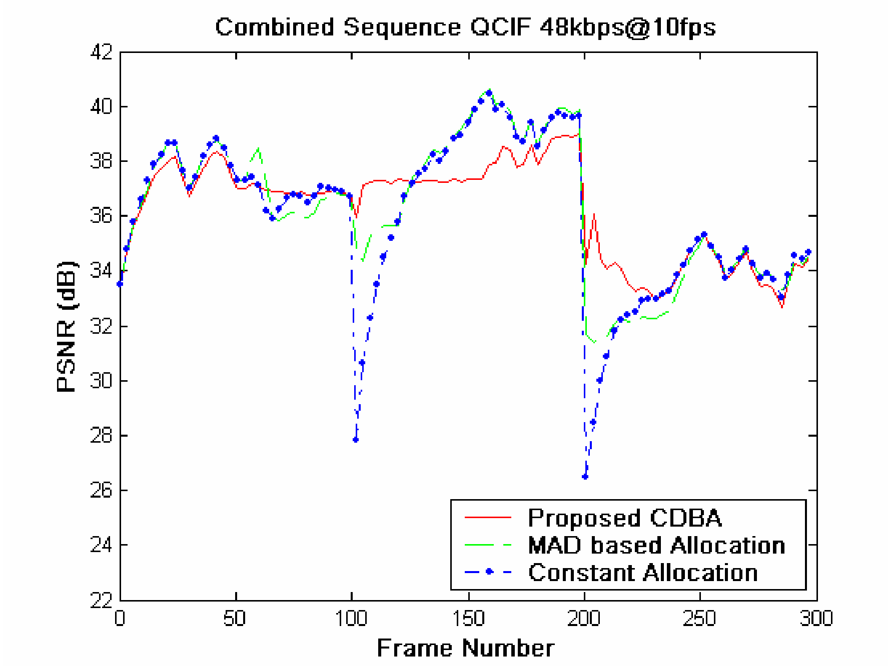
Mother & Daughter QCIF 48kbps@10fps
PSNR (dB)
Frame Number
Proposed CDBA
MAD based Allocation
Constant Allocation

(a) M&D



News QCIF 48kbps@10fps
PSNR (dB)
Frame Number
Proposed CDBA
MAD based Allocation
Constant Allocation

(b) News



(c) Combined Sequence of M&D, Akiyo and CarPhone

Figure 3. Frame by Frame PSNR comparison

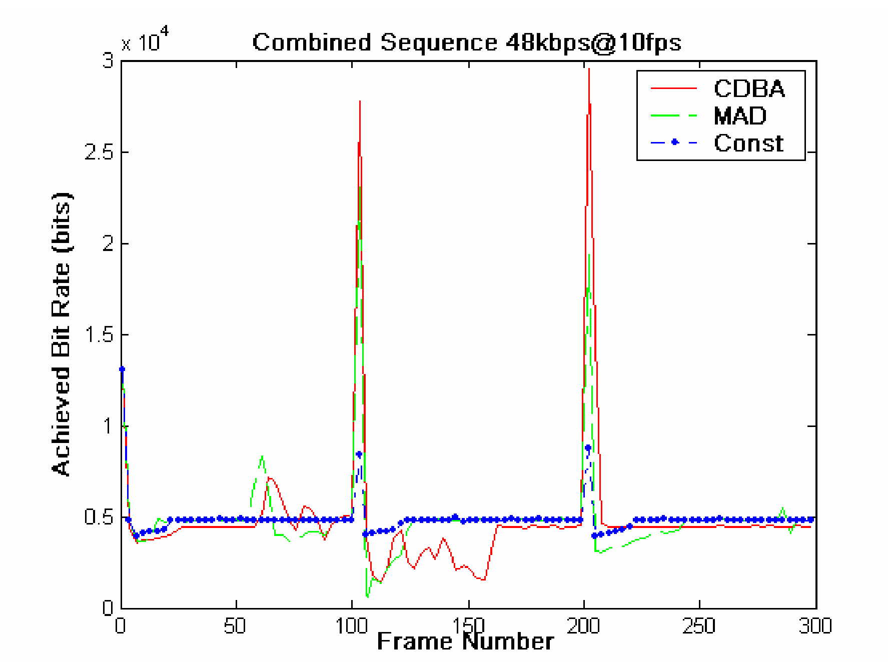


Figure 4. Bit Rate vs. frame number

(a) (b) (c)

(d) (e) (f)

Figure 5. Subjective quality comparison of the combined sequence. The 34th coded frame of the sequence: (a) CDBA (b) MAD based bit allocation and (c) constant bit allocation; The 67th frame of the sequence: (d) CDBA (e) MAD based bit allocation and (f) constant bit allocation

Table 1. Performance comparison between the Proposed CDBA, MAD based bit allocation and Constant Bit Allocation

Sequence	Algorithm	Actual bit rate (Kbps)	PSNR (dB)	σ^2 of PSNR
News	CDBA	48.48	33.96	0.67
	MAD	48.49	33.94	0.77
	Constant	48.48	33.93	0.97
M&D	CDBA	48.48	37.57	1.06
	MAD	48.48	37.60	1.68
	Constant	48.48	37.61	1.64
Foreman	CDBA	48.49	30.47	1.25
	MAD	48.48	30.37	2.60
	Constant	48.48	30.48	1.91
Combined	CDBA	48.48	36.23	3.20
	MAD	48.48	36.24	6.46
	Constant	48.48	36.00	8.38