Optimum Bit Allocation and Rate Control for H.264/AVC

Wu Yuan, Shouxun Lin, *Member, IEEE*, Yongdong Zhang, Wen Yuan, and Haiyong Luo

***Abstract*—For the rate control of H. 264/AVC, one of the most important things is to get the statistics of the current frame accurately. To achieve this, a novel adaptive coding characteristics prediction scheme is presented to improve the accuracy of R-D modeling, by exploiting spatio-temporal correlations. With the proposed prediction scheme, we present a novel rate function and a linear distortion model, and then deduce a simple close-form solution to the problem of optimum bit allocation, just in a TMN-8-alike way. Extensive experiments show that improvements with gains up to 0.92 dB per frame over JVT-G012, the current standardized rate control scheme, are achieved by the proposed scheme for a variety of test sequences with less demanding bandwidth.**

***Index Terms*—H.264/AVC, optimum bit allocation, rate control, rate-distortion optimization.**

I. Introduction

THE objective of rate control is to regulate the coded bit stream to satisfy certain given conditions, such as buffer over or underflow prevention as well as variable and/or low-bandwidth constraints.

As one of the key problems in regards to coding performance, rate control has drawn significant research attention, though it is currently left out of the specification scope of most hybrid video coding standards, such as MPEG-2[1], H.263[2], MPEG-4[3], and H.264/AVC[4].

A. Brief Reviews of Rate-Control Schemes

Generally speaking, a typical rate-control scheme consists of two basic operations: 1) bit allocation and 2) bit allocation achievement, namely bit rate control. The optimal rate control can be achieved jointly by optimum bit allocation and accurate bit-rate control. The task of optimum bit allocation is to efficiently distribute the bits budget among image blocks so that the best video quality is achieved. The problem of optimum bit allocation can be formulated as the following:

$$\min D, \qquad \text{subject to } R < R_c \tag{1}$$

Manuscript received March 28, 2005; revised October 12, 2005. This work was supported in part by the National Nature Science Foundation of China (60302028) and in part by the Key Project of International Science and Technology Cooperation (2005DFA11060). This paper was recommended by Associate Editor H. Sun.

Wu Yuan is with the Institute of Computing Technology, Chinese Academy of Sciences (CAS), Beijing 100080, China. He is also with the Graduate School, CAS (e-mail: wyuan@ict.ac.cn).

S. Lin, Y. Zhang, and H. Luo are with the Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China (e-mail: sxlin@ict.ac.cn;zhyd@ict.ac.cn; yhluo@ict.ac.cn).

Wen Yuan is with the Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China (e-mail: yuanw@lreis.ac.cn).

Digital Object Identifier 10.1109/TCSVT.2006.875215

where D denotes the overall distortion, R_c denotes the bit budget, and R denotes the number of bits used to encode the frame.

To achieve the target bit rate, the rate-control scheme appropriately chooses a quantization parameter. For accuracy, it is of importance to exactly model or estimate the coding bit rate in terms of the quantization parameter, namely rate-quantization $(R-Q)$ functions. Together with distortion-quantization $(D-Q)$ functions, $R-Q$ functions characterize the rate-distortion $(R-D)$ behavior of video encoding, which is the key issue of optimum bit allocation [25].

Many $R-Q$ and $D-Q$ functions have been reported in previous studies in [5]–[13], and various bit allocation schemes are simultaneously presented. Some of them have been adopted in various standard-compliant video coders, such as TM-5[12], TMN-8[11], and VM-8[13]. However, among these rate-control schemes, only TMN-8 achieves optimum bit allocation at a macroblock-by-macroblock basis.

B. Rate Control in H.264/AVC

H.264/AVC is a recently approved video-coding standard and achieves a significant improvement in coding performance in relation to prior coding standards at the price of increased complexity [14]. The Lagrangian coder control method [26] is incorporated into the H.264/AVC coder for high coding performance. However, the Lagrangian coder control method demands that the quantization parameter be evaluated before intra/inters prediction, thus leading to a dilemma because until the end of intra/inters prediction, the rate-control scheme cannot access the statistics, which is indispensable for the quantization parameter calculation. Such a dilemma prevents the rate-control scheme from directly accessing the statistics in advance.

In JVT-D030 [15], a two-pass scheme was proposed by Ma to circumvent the above dilemma, where a TM-5-alike method is used in each pass. If the first pass fails to obtain an appropriate quantization parameter, then a second pass will be conducted as a refinement, which consequently increases the computational complexity. Moreover, JVT-D030 uses an extremely simplified R-D function and, therefore, fails to achieve accurate and robust rate control.

In JVT-G012 [16], a one-pass scheme proposed by Li, spatio-temporal correlations are exploited to circumvent the dilemma, wherein a linear MAD model is employed to predict the coding complexity, and the conventional MPEG-4 Q2 function to calculate the quantization parameter. Nevertheless, spatio-temporal correlations are not sufficiently well utilized, consequently resulting in a weak prediction for the coding characteristic.

JVT-D030 and JVT-G012 both fail to achieve optimum bit allocation. In JVT-G012, the total target bits are distributed just proportionally to the coding complexity of each image block.

C. Proposed Optimum Bit Allocation and Bit-Rate Control Scheme for H.264/AVC

We aim to obtain a computationally feasible solution to the problem of optimum bit allocation for the H.264/AVC standard at a macroblock-by-macroblock basis. To achieve this goal, we should accurately model R-D behaviors and, therefore, first circumvent the obstacle introduced by the Lagrangian coder control method. As a way out, a novel adaptable prediction scheme is presented to accurately estimate the coding characteristic of the video content, wherein a spatio-temporal continuity criterion is employed both when data points are being selected for the regress analysis process and when coding characteristics are predicted by using the resulting R-D models.

After then, a linear distortion model and a modified Q2 rate model are presented and their accuracies are improved with the proposed prediction scheme. Based on them, we reduce a close-form solution to the problem of optimum bit allocation by using Lagrange optimization.

We implement our new rate-control scheme into JM-10[24], and compare its performance to that of the JVT-G012 rate-control scheme. The experimental results show that the proposed scheme performs the bit allocation more accurately and effectively. In comparison to the JVT-G012 rate control, the proposed scheme significantly increases the video quality up to 0.92 dB per frame.

D. Paper Organization

The rest of the paper is organized as follows. First, in Section II, a novel coding characteristics prediction scheme is presented. And then, we derive R-D models in Section III. In Section IV, we deduce a close-form solution to optimum bit allocation. After then, we propose a rate-distortion optimized rate control scheme in Section V. Extensive experiments are conducted to evaluate the performance of the proposed rate control scheme in Section VI. This paper concludes with Section VII.

II. Global Encoding Charicteristics Prediction

Generally speaking, the variations of the video content can be typically viewed as the results of movements of video objects relative to the imaging plane. The movements of real-world video objects intrinsically possess high spatio-temporal correlations/dependents, accordingly, which are also revealed in the video content. Thus, we can find great similarities available among consecutive pictures and among spatially adjacent blocks, which are almost removed by intra/inter predictions in the hybrid coding. Nevertheless, significant correlations are still available in the residual coding. In some works, spatio-temporal correlations have been exploited to speed up motion estimation [17]–[20] to predict the coding complexity [16]. In some works related to rate control, spatio-temporal correlations provide the insight into the assumption that neighboring blocks share a similar R-D curve [10], [16], [27].

However, the previous works related to rate control suffer from large prediction errors due to poor exploitations on correlations, as stated later, when predicting the coding characteristics, such as the overhead bit rate, the coding complexity, and the R-D behaviors, etc. As aforementioned, the performance of rate-control schemes depends heavily on R-D modeling. In order to improve R-D modeling, we present a heuristic adaptive coding characteristic prediction scheme in the following. It consists of two methods.

A. Improved Regress Method

Exploitations on spatio-temporal correlations with respect to rate control depend heavily on the linear regress method. How to select data points is very important to the linear regress analysis for a certain model. Generally speaking, improper exploitations on correlations often lead to the bad selection of data points and, therefore, inaccurate R-D modeling and, consequently, leading to large prediction errors. Unfortunately, the previous works just do in an imperfective way for which we list three primary deficiencies together with the proposed improvements in the following, respectively.

1) Exclusively Exploiting Spatial and Temporal Correlations: In the regress analysis for the MPEG-4 Q2 model and the linear MAD model, data points are selected from recently coded blocks, which are situated in the same frame except at the startup time. In such a case, are only spatial correlations used. The previous works often suffer from relative large prediction errors in case of low spatial correlations available.

Improvement: Exploit Both Spatial and Temporal Correlations: Generally speaking, temporal correlations are ubiquitous throughout the whole video content. It is easy to see that jointly exploiting spatial and temporal correlations is more adaptive and robust than exploiting only spatial correlations when selecting data points. Furthermore, the often-occurring scenario that no spatial but only temporal correlations are available necessitates exploitations on temporal correlations.

2) Poor Data Points Selection: The quality and quantity of the data set used for regress analysis are critical and should qualify at least enough to differentially illustrate a variety of spatio-temporal continuities. For the MPEG-4 Q2-rate model and the linear MAD model, data points are simply collected along the raster scanning path. Such a data-point selection scheme is too simple to comply with the actual spatial continuity. There is the risk that data points with low correlations are involved and, therefore, weaken the accuracy of R-D models.

Improvement: Select Data Points According to Spatio-Temporal Continuities: Generally, the video content with the same spatio-temporal continuity is irregularly shaped. It is very difficult to exactly segment the video content with respect to spatio-temporal continuities. Observe that spatio-temporal correlations between blocks strongly depend on their spatial and temporal distance. The closer the distance is, the higher the spatio-temporal correlations are, and vice-versa. Thus, a heuristic simple solution is presented, which is similar to the slide widow technique [21] with the only difference being the preliminary stage of data-point selection. At the preliminary stage, data points are collected from only these coded macroblocks within a certain spatial distance in the current and previous P frames in the proposed prediction scheme, while data points are collected from recently coded macroblocks in [21]. A refinement is conducted

against the preliminary data set by removing all of the outliner data points in the same way as the slide widow method does. Note that there may be little spatial correlations for spatially remote macroblocks at the preliminary stage. In such cases, only temporal correlations are used.

3) Misuse of R-D Models in Prediction: Via linear regress analysis, we can achieve a certain R-D model which consists of the statistics of the data set. Thus, a model is temporally and spatially volatilizable, and we can use it to estimate the coding characteristics only for macroblocks of the same spatio-temporal continuity. Applying the resulting R-D models without good compliancy to spatio-temporal continuity will lead to significant prediction errors. Unfortunately, JVT-G012 has the resulting MAD model to be used in predictions for all macroblocks, even for spatially remote ones which may have low spatial correlations, thus leading to large prediction error.

Improvement: Properly Use the Resulting R-D Model: For not yet coded macroblocks, we cannot be certain what spatio-temporal continuity they are in. Nevertheless, it is observed that spatially or temporally adjacent macroblocks are highly correlated at a high probability. Thus, a feasible but not perfective method is proposed, wherein the resulting R-D models are used in prediction only for spatially and temporally adjacent macroblocks.

B. Simple Prediction Method

It is very hard to exactly model the overhead bit rate and the distortion. It is stated in [17] that the motion vector (MV) of a certain block may be very close to that of its temporally adjacent blocks. It hints that the overhead bit rate of a certain macroblock may also be very close to that of its temporally adjacent macroblocks, since MV spends most of the overhead bits. It is justified by experiments, as stated later. Similar observations can be achieved with respect to the distortions. For simplicity, we can use the temporal history as a substitute or a scaling factor when predicting the overhead bit rate and the distortions. A detailed description of this method is provided in the subsequent section.

III. Rate-Distortion Modeling

As aforementioned, the problem of R-D modeling is complicated for H.264/AVC due to its increased complexity. For the prior coding standards, studies are mainly concentrated on R-Q and D-Q functions, while for H.264/AVC, additional attention should be paid to how to get the statistics of the current frame. With a sight into the problem of R-D modeling, we decompose it into four substantial components, including overhead bit-rate prediction, coding complexity prediction, R-Q modeling and D-Q modeling, which will be described in detail in the following, respectively.

A. Overhead Bit-Rate Prediction

The overall bitstream generated by the source coder is mainly comprised of bits used for texture coding plus bits for differentially coding overhead information, such as the macroblock mode, the motion vector, and the quantization parameter. Most of prior rate-control schemes regard the overhead bits as a constant component in the overall bitstream, and update it with the average overhead bit rate after coding a macroblock or a frame.

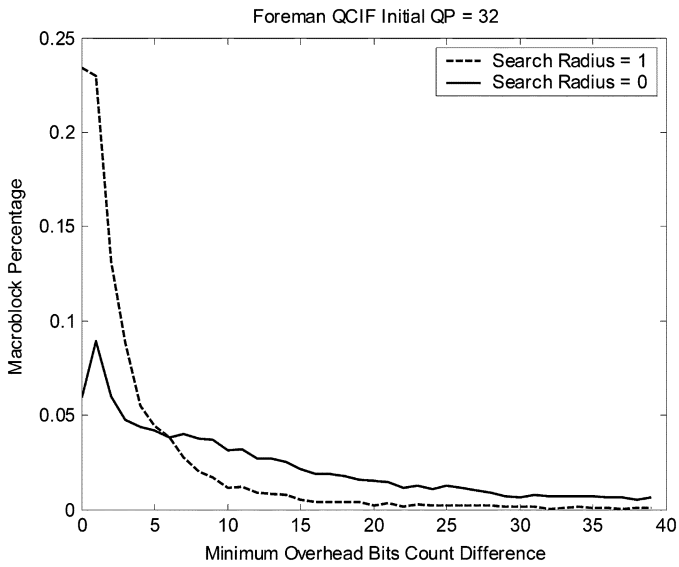


Fig. 1. Temporal correlation of overhead bits count.

The conventional method does not fit H.264/AVC anymore, because the coding structure has become complicated, leading to a possible dramatic change in the number of bits required to code the overhead information. Unfortunately, the JVT-G012 rate-control scheme follows that way and is vulnerable to large prediction error.

In this work, exploitations of spatio-temporal correlations are conducted to improve the accuracy of the overhead bit-rate prediction. As previously mentioned, extensive correlations exist among motion vectors of spatially and temporal adjacent macroblocks [17]. Furthermore, it can be also observed that similar mode selections are preferred among spatially and temporal adjacent macroblocks. All of these observations strongly suggest that the overhead bit rates may be close among spatially and temporally adjacent macroblocks. Experiments[1] are conducted to justify this, and the typical experiment results are plotted in Fig. 1. The experiment is done by searching for the minimum difference of the overhead bits count between the current macroblock and all of the macroblocks in the previous P frame with a distance to the co-located position of the current macroblock not exceeding the searching radius.

Therefore, we can use the overhead bit rate of the previous P frame to predict that of the current P frame. It is shown in Fig. 1 that the prediction with the searching radius equaling 1 is better than that with the searching radius equaling 0. However if the searching radius is other than 0, we will get confused with the choice among more than one macroblocks. So, for simplicity, we propose to predict the overhead bits count of not yet coded macroblocks directly by that at the colocated position in the previous P frame. It is formulated as the following:

$$H_i = H_i^{\mathrm{prev}} \tag{2}$$

where H_i denotes the overhead bit rate of the ith macroblock in the current P frame, and H_i^{prev} denotes the overhead bit rate of the ith macroblock in the previous P frame.

[1]In this work, all of the experiments are done under the same test conditions as stated in Section VI.

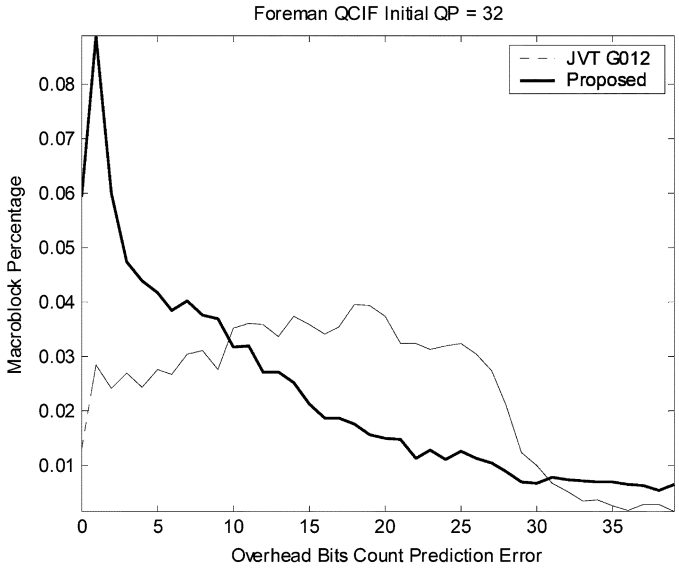


Fig. 2. Overhead bit-rate prediction comparison.

Extensive experiments have been done to verify the proposed method, and the typical results are presented in Fig. 2. It is shown that the new formula is more accurate and robust than JVT-G012.

B. Coding Complexity Prediction

Coding complexity reflects the substantial of the video content and is a valuable tool used to predict the texture bit count prior to encoding a frame or macroblock. Various statistics measures can be used as indications of coding complexity, such as MAD and mean square error (MSE).

In the prior video-coding standards, the actual coding complexity is accessible immediately after intra/inters prediction, and can be used in the subsequent rate-control process. However, it is not the case for H.264/AVC, because the Lagrangian coder control method desires the rate-control operation to occur in advance. To circumvent the obstacle, a linear MAD model is proposed in JVT-G012, which attempts to approximate the temporal variation of coding complexity. With the linear MAD model, MAD of not yet coded macroblocks can be predicted by that in the colocated position in the previous P frame. The linear MAD model is formulated by the following:

$$\mathrm{MAD} = \rho \mathrm{MAD}^{\mathrm{prev}} + \gamma \tag{3}$$

where MAD denotes the predicted MAD of the current macroblock, $\mathrm{MAD}^{\mathrm{prev}}$ denotes the actual MAD of the macroblock in the colocated position of the previous P frame, and ρ and γ denote the two model parameters.

However, as previously stated, the prediction performance of the linear MAD model is weakened by the poor use of spatio-temporal correlations in linear regress analysis. We improve it by applying the proposed prediction scheme. Experiments are conducted to justify it and the performance evaluation is reported in Table I, where the first column shows the relative prediction error (RPE) defined by $|\mathrm{MAD}_{\mathrm{predicted}} - \mathrm{MAD}_{\mathrm{actual}}|/\mathrm{MAD}_{\mathrm{predicted}} \times 100\%$, and the second to fourth columns indicate the percentage of macroblocks related to the RPE. We can see that the proposed method is a little better than that of JVT-G012.

TABLE I
MAD PREDICTION ERROR COMPARISON

Relative Prediction Error	Foreman Initial QP=32		News Initial QP=34	
	JVT-G012	Proposed	JVT-G012	Proposed
0.00%	5.34%	5.68%	26.83%	29.25%
1.00%	5.68%	6.54%	8.96%	13.88%
2.00%	4.81%	5.62%	6.36%	8.61%
3.00%	4.75%	5.21%	5.83%	6.38%
4.00%	4.64%	5.50%	4.51%	4.87%
5.00%	4.16%	5.38%	3.26%	3.58%
6.00%	3.89%	4.95%	2.89%	3.17%
7.00%	4.36%	4.13%	2.75%	3.06%
...	...	...	...	...

C. R-D Behavior Prediction

It is a well-known assumption that the discrete cosine transform (DCT) coefficients of the motion-compensated difference frame are approximately uncorrelated and Laplacian distributed alike (4)

$$P(x) = \frac{e^{-\frac{|x|}{\sigma}}}{2\sigma} \quad \text{where } -\infty < x < \infty. \tag{4}$$

If the distortion measure is defined as $D(x, \bar{x}) = |x - \bar{x}|$, we can get a closed-form solution for the R-D functions as (5), according to mathematics derivation in [22]

$$R(D) = \ln(\frac{\sigma}{D}) \quad \text{where } 0 < D < \sigma. \tag{5}$$

We assume the quantization step size Qstep as the distortion measure. By using Taylor expansion, we obtain a formula as follows:

$$\begin{aligned} R(\mathrm{Qstep}) &= \left(\frac{\sigma}{\mathrm{Qstep}} - 1\right) - \frac{1}{2}\left(\frac{\sigma}{\mathrm{Qstep}} - 1\right)^2 \\ &\quad + R_3(\mathrm{Qstep}) \\ &= -\frac{3}{2} + \frac{2\sigma}{\mathrm{Qstep}} - \frac{\sigma^2}{2\mathrm{Qstep}^2} + R_3(\mathrm{Qstep}). \end{aligned} \tag{6}$$

By neglecting $R_3(\mathrm{Qstep})$ and replacing σ with MAD, we obtain a simple quadric formula as follows:

$$R(\mathrm{Qstep}) = \frac{a\mathrm{MAD}}{\mathrm{Qstep}} + \frac{b\mathrm{MAD}^2}{\mathrm{Qstep}^2} \tag{7}$$

where the two parameters a and b are calculated using the linear regression method. The proposed prediction scheme is applied when data points are being selected in the regress analysis process.

We find it very difficult to obtain a simple close-form solution to the problem of optimum bit allocation via (7), which motives us to look for a new model. Generally speaking, in order to maintain the smoothness of visual quality, the admittable quantization parameter in a frame is restricted to a narrow range and, meanwhile, the quantization parameter between adjacent macroblocks cannot be changed rapidly and abruptly. We assume that the admittable quantization parameter in the current P frame varies from $QP_{\min}$ to $QP_{\max}$, and QP_{ave} is the midpoint. $\text{Qstep}_{\min}, \text{Qstep}_{\max}$ and $\text{Qstep}_{\text{ave}}$ are the quantization step sizes of $QP_{\min}, QP_{\max}$ and QP_{ave}, respectively. Then, we expand (7) into a Taylor series at the point of $\text{Qstep}_{\text{ave}}$ and obtain a formula as follows:

$$\begin{aligned}\tilde{R}(\text{Qstep}) = {} & R(\text{Qstep}_{\text{ave}}) + R'(Q\text{step}_{\text{ave}}) \\ & \times (\text{Qstep} - \text{Qstep}_{\text{ave}}) \\ & + \frac{R''(\text{Qstep}_{\text{ave}})}{2} \times (\text{Qstep} - \text{Qstep}_{\text{ave}})^2 \\ & + R_3(\text{Qstep} - \text{Qstep}_{\text{ave}}) \end{aligned} \tag{8}$$

where $R'(\text{Qstep}_{\text{ave}})$ and $R''(\text{Qstep}_{\text{ave}})$ are the first and second derivative of (7) at $\text{Qstep}_{\text{ave}}$, respectively. Considering that Qstep is confined in the range of $[QP_{\min}, QP_{\max}]$, we neglect $R_3(\text{Qstep} - \text{Qstep}_{\text{ave}})$ in (8), and then we can achieve (9). Though (9) maybe suffers from a large truncation error in case that Qstep deviates a lot from $\text{Qstep}_{\text{ave}}$, the experimental results in Section VI prove it a pragmatically feasible approach

$$\tilde{R}(\text{Qstep}) = A \times \text{Qstep}^2 + B \times \text{Qstep} + C \tag{9}$$

where

$$\begin{aligned} & \text{Qstep}_{\min} < \text{Qstep}_i < \text{Qstep}_{\max}, \\ & A = \frac{R''(\text{Qstep}_{\text{ave}})}{2}, \\ & B = R'(\text{Qstep}_{\text{ave}}) - R''(\text{Qstep}_{\text{ave}}) \times \text{Qstep}_{\text{ave}}, \end{aligned}$$

and

$$C = R(\text{Qstep}_{\text{ave}}) - A \times \text{Qstep}_{\text{ave}}^2 - B \times \text{Qstep}_{\text{ave}}.$$

Combining (2) and (9), we can obtain a rate model for the ith macroblock presented in the following:

$$R_i = A_i \text{Qstep}_i^2 + B_i \text{Qstep}_i + C_i + H_i^{\text{prev}} \tag{10}$$

where

$$\text{Qstep}_{\min} < \text{Qstep}_i < \text{Qstep}_{\max}.$$

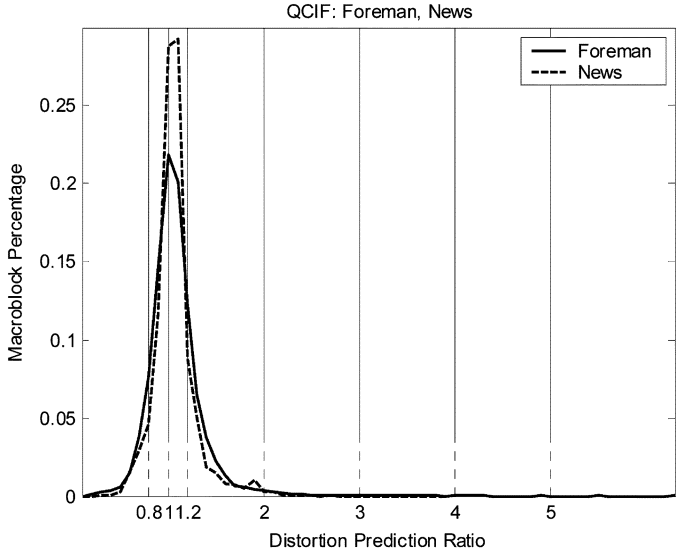


Fig. 3. Verification of the proposed distortion model.

D. Distortion Prediction

The distortion between the reconstruction frame and the original one is introduced by uniformly quantizing DCT coefficients. Some distortion models, such as MAD, MSD, or peak signal-to-noise ratio (PSNR) are used in actual comparisons.

It is shown that MSD is used as the distortion measurement [6], [22], [27]. Under the assumption that the distortion of DCT coefficients is uniformity distributed, we can then achieve a formula of $D = (\text{Qstep}^2/12)$ to model the distortion. Nevertheless, this model is not scalable with video contents.

We observe that the distortion is distributed similarly in spatially and temporally adjacent macroblocks mostly due to spatio-temporal correlations. Heuristically, we can add the history into the distortion model as an index of the video content. With the assumption of $\text{Qstep}_i \propto \text{Qstep}_i^{\text{prev}}$, we can see $\text{Qstep}_i \propto (D_i^{\text{prev}}/\text{Qstep}_i^{\text{prev}})$ and further $D_i \propto \text{Qstep}_i^2 \propto (\text{Qstep}_i D_i^{\text{prev}}/\text{Qstep}_i^{\text{prev}})$. Thus, we can accomplish a heuristic linear distortion model presented as (11)

$$D_i = \zeta \times \chi_i \times \text{Qstep}_i, \text{ where } \chi_i = \frac{D_i^{\text{prev}}}{\text{Qstep}_i^{\text{prev}}} \tag{11}$$

where ζ is a constant which, we can see later, is not required in the computation for optimum bit allocation.

Extensive experiments have been done to verify the proposed distortion model, and typical test results are depicted in Fig. 3 with the distortion prediction ratio defined as $D_{\text{actual}}/D_{\text{predicted}}$, where D_{actual} is the actual distortion and $D_{\text{predicted}}$ the predicted distortion. We can see in Fig. 3 that the value of ζ is about 1.0. If ζ is set to 1.0, the percentage of macroblocks with an average prediction error (APE), defined as $|D_{\text{predicted}} - D_{\text{actual}}|/D_{\text{predicted}}$, below 10% is 42% for foreman and 58.1% for news, and the percentage with APE below 20% is 68.8% for foreman and 78.6% for news.

IV. Rate Distortion Optimization

Based on the above R-D models, a mathematics closed-form solution to the R-D optimized rate control is addressed in the following. We want to find an expression for the quantization parameters that minimizes the distortion in (11) subject to the constraint that the total number of bits [i.e., the sum of the macroblock's (10)], must be equal to the frame target budget T

$$\begin{aligned}&\mathrm{Qstep}_1^*,\ldots,\mathrm{Qstep}_N^*\\&\quad=\underset{\substack{\mathrm{Qstep}_1,\ldots,\mathrm{Qstep}_N\\ \sum_{i=1}^N T_i=T}}{\arg\min}\ \frac{1}{N}\sum_{i=1}^{N}\zeta\times\chi_i\times\mathrm{Qstep}_i.\end{aligned}\quad(12)$$

Since we are minimizing a convex, differentiable function on a convex set, there is a unique solution that can be obtained using Lagrange theory [23]. To do this, we define λ and λ^* as, respectively, the Lagrange multiplier and its optimal value, and express the optimization problem in (12) in its equivalent form

$$\begin{aligned}&\mathrm{Qstep}_1^*,\ldots,\mathrm{Qstep}_N^*,\lambda^*\\&=\underset{Q_1,\ldots,Q_N,\lambda}{\arg\min}\ \frac{1}{N}\sum_{i=1}^{N}\zeta\chi_i\mathrm{Qstep}_i+\lambda\left[\sum_{i=1}^{N}T_i-T\right]\\&=\underset{Q_1,\ldots,Q_N,\lambda}{\arg\min}\ \frac{1}{N}\sum_{i=1}^{N}\zeta\chi_i\mathrm{Qstep}_i\\&\quad+\lambda\left[\sum_{i=1}^{N}\left(A_i\mathrm{Qstep}_i^2+B_i\mathrm{Qstep}_i+C_i+H_i^{\mathrm{prev}}\right)-T\right]\end{aligned}\quad(13)$$

where we replace T with (10) in the last step.[2]

After some straightforward manipulations, we obtain the optimal quantizer step in (14). It is observed that ζ is dissolved in (14)

Equation (14), shown at the bottom of the page, is the key formula in the novel proposed rate-control scheme. Note that we cannot achieve a numerical solution to (14) under two conditions.

Condition 1) The formula (10) falls back to a first-order model if $A_i = 0$, and even a constant model if $A_i = 0$ and $B_i = 0$. For the former case, we can conduct an adjustment on the model parameters A_i, B_i and C_i by fitting the linear model with a quadratic curve.

Condition 2: If

$$\frac{T-\sum_{k=1}^{N}H_k^{\mathrm{prev}}+\sum_{k=1}^{N}\frac{B_k^2}{4A_k}-\sum_{k=1}^{N}C_k}{\sum_{k=1}^{N}\frac{\chi_k^2}{A_i}}<0$$

the calculation on (14) leads to a complex solution, it is because that the frame target is too small or too large.

[2]Here, the constraint of $\mathrm{Qstep}_{\min}<\mathrm{Qstep}_i<\mathrm{Qstep}_{\max}$ on the formula (10) is removed just to simplify the deduction. Extensive experiments in Section VI show it pragmatically feasible.

V. Rate Control

We aim to design a rate-control scheme at the macroblock layer and follow the way in JVT-G012 when allocating the bits quota at the group of pictures (GOP) level and frame level. In JVT-G012, a fluid traffic model based on the linear tracking theory is employed to estimate target bits for the current GOP and frame, and the detailed information can be referred to in [13].

A. Macroblock-Layer Rate Control

We follow the way of JVT-G012 to determine the starting quantization parameter QP_{start} of a GOP, which is used to encode I frame and the first P frame. For the remaining P frames, the quantization step sizes are optimally evaluated for all of the macroblocks so that the actual number of encoded bits is close to the target bit-budget T. The following is a step-by-step description of the method.

Step 1: Initialization: If the current P frame is the second P frame in the current GOP, let $QP_{\mathrm{ave}} = QP_{\mathrm{start}}$. Compute $QP_{\min} = \max(QP_{\mathrm{ave}} - \Delta, 0), QP_{\max} = \min(Q\mathrm{Pave} + \Delta, 51)$, where Δ equals 3. Compute the quantization step size $\mathrm{Qstep}_{\mathrm{ave}}, \mathrm{Qstep}_{\min}$ and $\mathrm{Qstep}_{\max}$ related to $QP_{\mathrm{ave}}, QP_{\min}$ and $QP_{\max}$, respectively.

Let $i = 0$.

Step 2: Optimum Bit Allocation for the ith Macroblock:

Step 2.1: If the current frame is the first P frame in the current GOP, QP^* is set to QP_{start}. Then, it jumps to ***Step 4***.

Step 2.2: Otherwise, let $k = i, \mathrm{SD2\_A} = 0, \mathrm{SC} = 0, \mathrm{SB2\_A} = 0, \mathrm{SO} = 0$ and $\mathrm{SL} = 0$; and then do the following iteration.

Step 2.2.1: R-D modeling

Compute a_k and b_k for the MPEG-4 Q2 model and β_k and γ_k for the linear MAD model by using

$$\begin{aligned}Q_i^*=&-\frac{B_i}{2A_i}-\frac{\chi_i}{A_i}\\&\times\sqrt{\frac{T-\sum_{k=1}^{N}H_k^{\mathrm{prev}}+\sum_{k=1}^{N}\frac{B_k^2}{4A_k}-\sum_{k=1}^{N}C_k}{\sum_{k=1}^{N}\frac{\chi_k^2}{A_i}}},\qquad i=1,\ldots,N.\end{aligned}\quad(14)$$

the linear regress method. Note that data points are first selected in regards to the spatial and temporal distance, and then outliers are removed in the regress analysis process.

Compute MAD_k.

Compute A_k, B_k and C_k for the proposed Q2 model according to (9).

Step 2.2.2: Adjustment on the proposed Q2-rate models

If $A_k = 0$ and $B_k \neq 0$, an adjustment[3] on the proposed Q2 rate models is conducted to leave $A_k \neq 0$.

Step 2.2.3: Optimum Computations

If $A_k = 0$, then let $SD2\_A = 0$ and $SB2\_A = 0$.

Otherwise, let $SD2\_A = SD2\_A + (\chi_k^2/A_i)$ and $SB2\_A = SB2\_A + (B_i^2/A_i)$.

Let $SC = SC + C_k$, $SO = SO + H_k$ and $SL = SL + (a_k \mathrm{MAD}_k/\mathrm{Qstep}_{\max}) + (b_k \mathrm{MAD}_k^2/\mathrm{Qstep}_{\max}^2)$.

Step 2.2.4: Loop Condition

Let $k = k + 1$.

If $k > N$, terminate the iteration and jump to Step 2.3.

Otherwise, jump back to *Step 2.2.1*.

Step 2.3: Compute Optimal QP_i^*

There are three cases we need to deal with.

In case that $A_i = 0$:

If $SO + SL > T$, let $QP_i^* = QP_{\max}$.

Otherwise, let $QP_i^* = QP_{\min}$.

Then, jump to Step 3.

In case that $(4T - 4SO - 4SC + SB2\_A/SD2\_A) < 0$:

If $SO + SL > T$, let $QP_i^* = QP_{\max}$.

Otherwise, let $QP_i^* = QP_{\min}$.

Then, jump to Step 3.

In the other case:

Compute the optimal quantization step size Qstep_i^* for the current macroblock using the following formula:

$\mathrm{Qstep}_i = -(B_i/2A_i) - (\chi_i/2A_i)\sqrt{(4T - 4SO - 4SC + SB2\_A/SD2\_A)}$.

And then deduce QP_i^* from Qstep_i^*.

Step 3: Adjust QP_i^*: Let $QP_i^* = \max\{QP_{i-1}^* - 1, \min\{QP_i^*, QP_{i-1}^* + 1\}\}$ in order to reduce the blocking artifacts. Then, it is further bounded by $QP_i^* = \max\{1, QP_{\mathrm{ave}} - \Delta, \min\{51, QP_{\mathrm{ave}} + \Delta, QP_i^*\}\}$ to maintain the smoothness of visual quality.

Step 4: Macroblock Encoding

Step 5: Post-Encoding: Update the coding history including the overhead bit rate, the distortion ratio $\chi_i = (SSD_i/\mathrm{Qstep}_i), X_i = (\mathrm{MAD}_i/\mathrm{Qstep}_i)$, and $Y_i = (R_i \mathrm{MAD}_i/\mathrm{Qstep}_i)$, where R_i is the texture bit rate. If the frame is the first P frame of the current GOP, then let $\mathrm{MAD}_i^{\mathrm{cur}} = \mathrm{MAD}_i$ and $\mathrm{MAD}_i^{\mathrm{prev}} = \mathrm{MAD}_i$; otherwise, let $\mathrm{MAD}_i^{\mathrm{prev}} = \mathrm{MAD}_i^{\mathrm{cur}}$ and $\mathrm{MAD}_i^{\mathrm{cur}} = \mathrm{MAD}_i$.

Step 6: Loop Condition: Let $i = i + 1$.

[3]The simple adjustment on the Q2 model is to design a quadric model to approximate the linear model, such that the two curves intersect at the two points with the quantization step size to equal $\mathrm{Qstep}_{\max}$ and $\mathrm{Qstep}_{\min}$, respectively, and the distance of points with the quantization step size between $[\mathrm{Qstep}_{\max}, \mathrm{Qstep}_{\min}]$ in the quadric model to the linear model does not exceed a threshold, which is defined as 1 in the subsequent experiment.

TABLE II
TEST CONDITIONS

MV resolution	1/4 pel	**Reference Frames**	1
Hadamard	ON	**Symbol Mode**	CABAC
RD optimization	ON	**GOP structure**	IPPP
Search Range	±32	**IntraPeriod**	0
Restrict Search Range	2	**Basic Unit**	1

If i exceeds the total number of macroblocks in the current frame, the encoding process for the current frame comes to an end; otherwise, go back to Step 2.

B. Computational and Memory Complexity Analysis

The proposed rate control proceeds over macroblocks in one-pass, and its computational complexity is comparative to that of JVT-G012. The computational complexity can be reduced by only selecting data points from the history data set of the previous P frame at the sacrifice of spatial correlation. Another approach to speeding up computation is to borrow the tool of the basic unit from JVT-G012.

Additional memory should be allocated for the storage of coding history, including two arrays for Q2 function, two arrays for the linear MAD model, one array for the distortion model, and one array for the overhead model. The total memory size is proportional to the total number of macroblocks in a frame, and is acceptable for a real coder system.

VI. EXPERIMENTAL RESULTS

Numerous experiments have been conducted to evaluate the performance of the proposed rate-control scheme with JVT reference software JM 10 [24] serving as a test benchmark. For fair play, the proposed rate-control scheme is implemented also based on JM 10, so that all parts of the execution binary except the rate-control module are the same.

Test sequences are Container, Foreman, Mobile, News, and Paris. Container, Foreman and News are in format of QCIF (4:2:0) while the rest are in format of CIF (4:2:0). The frame rate is 10 f/s for Container, Foreman, and News; 15 f/s for Paris; and 30 f/s for Mobile. The number of frames to be encoded is 100 for Container, Foreman, and news; 150 for Paris; and 300 for Mobile. The other test conditions are listed in Table II.

To achieve an evaluation of the proposed scheme in the rate-distortion sense, experiments are conducted at various bit rates. The test sequences are first encoded with fixed quantization parameters, which are 24, 26, and $28\cdots 44$, respectively, and then, the generated bit rates are used as the target bit rates when encoding using the proposed rate-control scheme and JM 10. The initial quantization parameters of the proposed rate-control scheme and JM 10 are set to the value of the fixed quantization parameters with respect to the target bit rates.

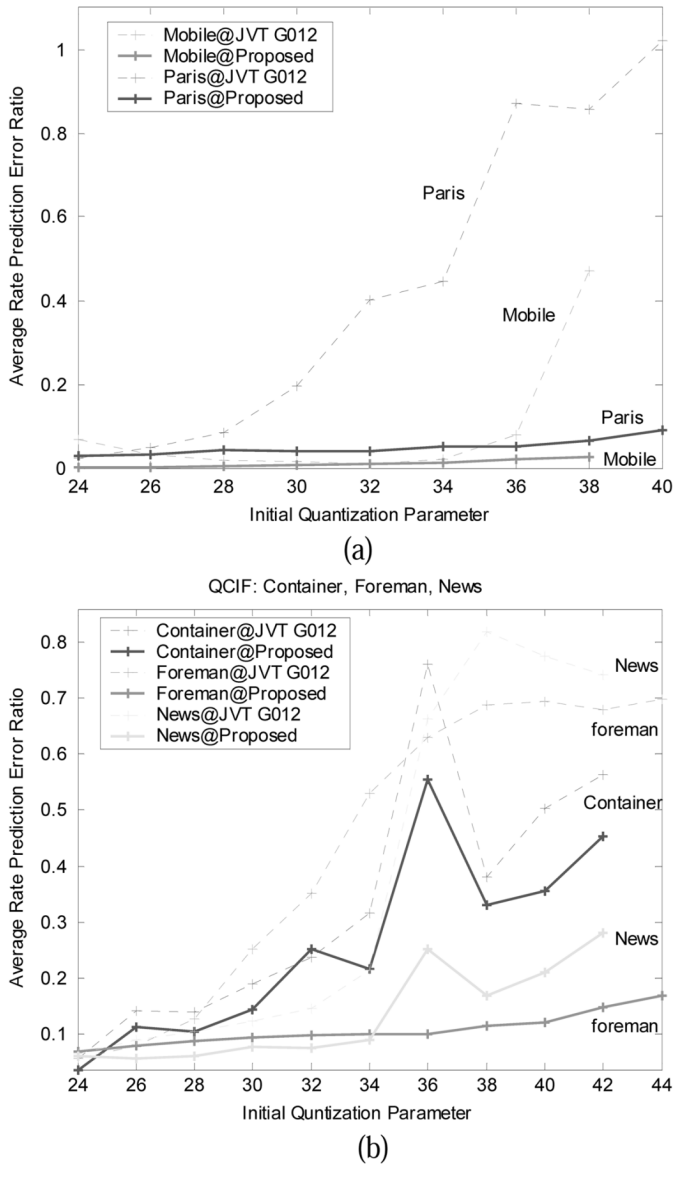


Fig. 4. Average rate prediction error ratio comparisons at the GOP level.

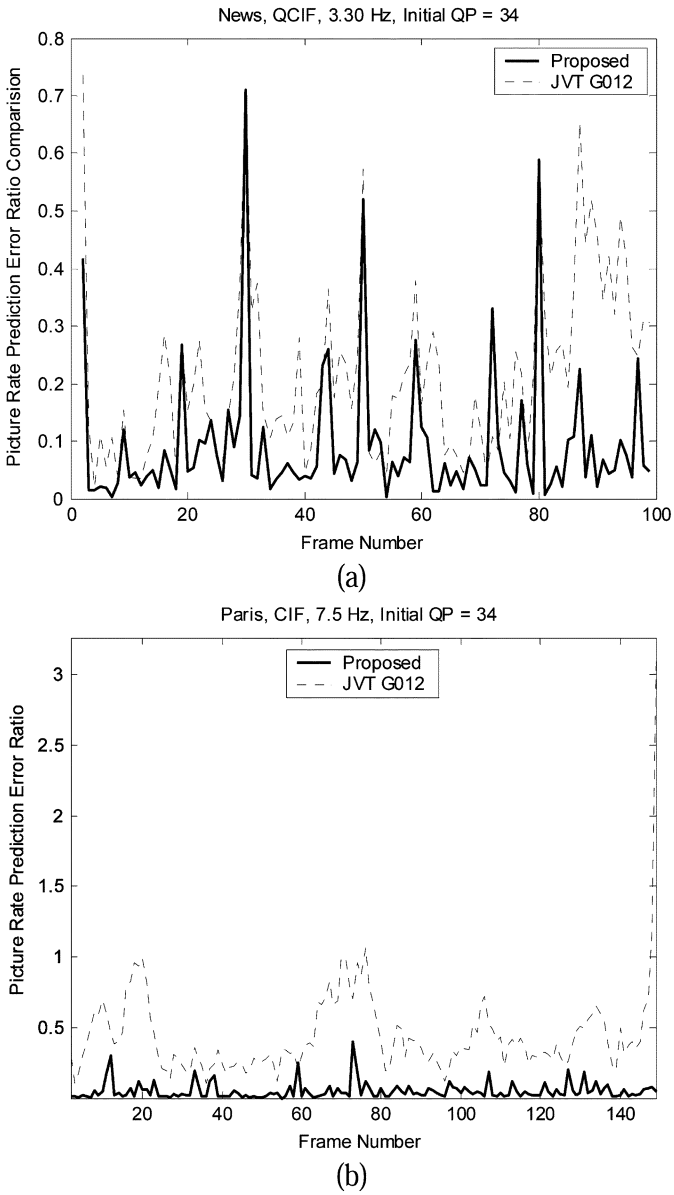


Fig. 5. Rate prediction error ratio at picture-level comparisons.

In nearly every compression application, achieving the target bit rate is of paramount importance. We will compare the performance of the proposed scheme to the JVT-G012 rate-control scheme, in terms of how effective each scheme achieves the target bit-budget for each picture. To quantify this performance, we use the rate prediction error ratio (RPER), defined as $|(R_a - T_p)/T_p| \times 100\%$, where T_p is the target picture bit budget and R_a is the actual number of encoded bits for the picture. A smaller RPER indicates superior scheme performance and vice-versa.

Fig. 4 shows the average RPERs at the GOP level for all of the tested initial quantization parameters, for all of the test sequences, and for the proposed scheme and JVT-G012. We can see that the average RPERs in the proposed scheme are much lower than that in JVT-G012 for all of the test sequences. And it is also shown in Fig. 4 that the average RPERs vary steadily and slowly for the proposed scheme for all of the sequences except Container, while it varies abruptly for JVT-G012, with respect to the initial quantization parameter. Further insight into the RPERs at picture level can be obtained from Fig. 5, where some typical test results are presented. All of these plots clearly show that the proposed scheme does a better job than JVT-G012 when choosing the quantization parameters, especially in the case of low bit rate, such that the target budget for a picture is achieved. All of these are evidence that the proposed scheme is more robust with respect to each video sequence. This robustness is probably due, at least in part, to the proposed coding characteristics prediction scheme and optimum bits allocation in the proposed rate-control scheme.

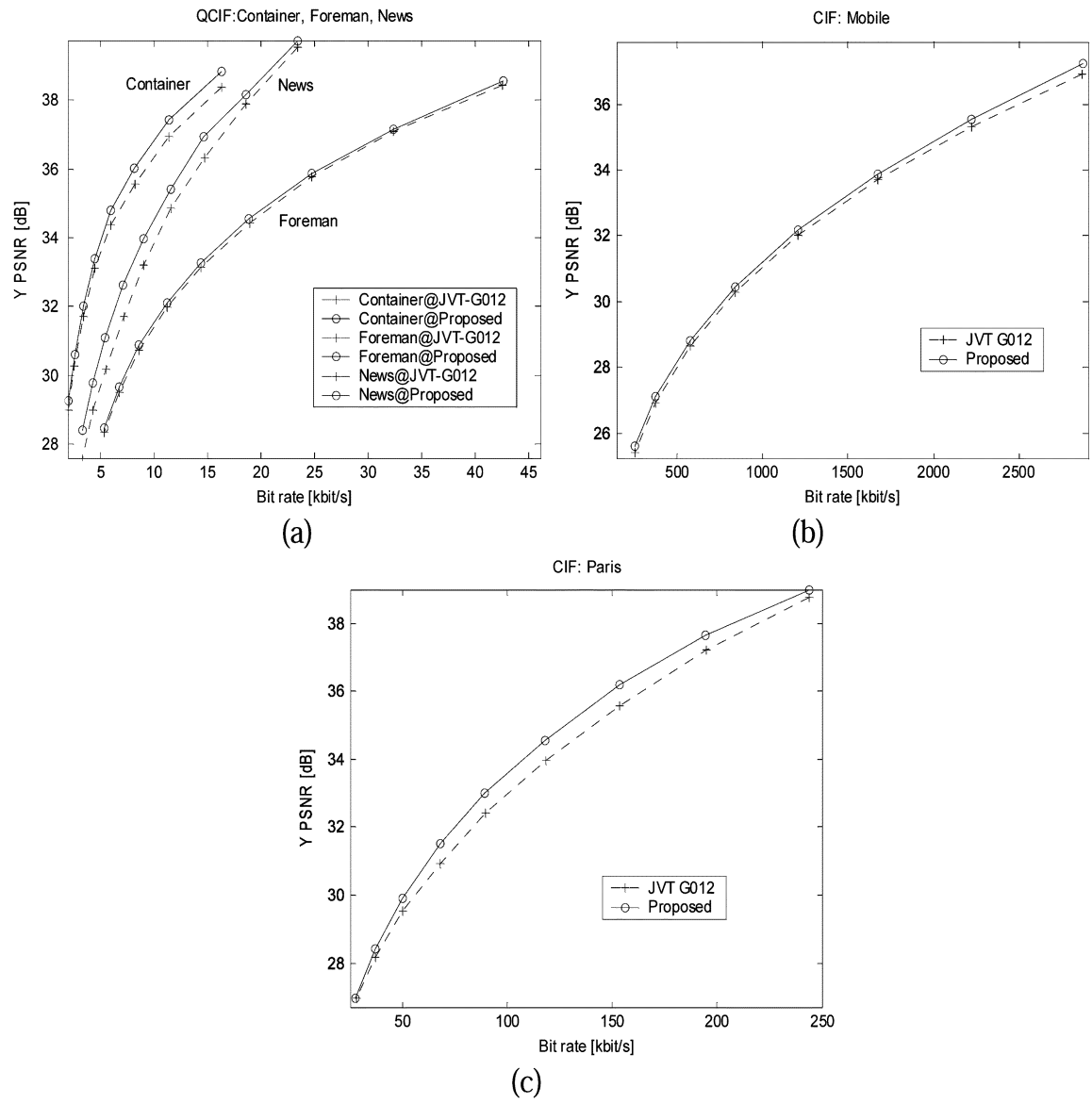


Fig. 6. Rate-distortion performance comparisons.

Besides controlling the number of bits spent in a macroblock, the quantization parameter also determines the distortion caused by quantization in a macroblock. In Fig. 6, we compare the overall rate-distortion performance of the proposed rate-control scheme with that of JVT-G012. It can be seen that the rate-distortion curves for the proposed scheme lie above that for JVT-G012 at almost all of the points for all of the test sequences. Note that the rate-distortion curves of Foreman for the proposed scheme are superior to the ones for JVT-G012, though they seem somewhat identical, respectively. For easy comparison, the average PSNRs of all of the test sequences and the gain thereof are also tabulated in Table III. It is demonstrated that the proposed rate-control scheme achieves higher coding performance in comparison with JVT-G012 with the gain up to 0.92 dB per frame and, meanwhile, the proposed rate-control scheme gains a mild save of the average bandwidth. The average PSNR gain of the proposed scheme is 0.323 dB for Container, 0.1 dB for Foreman, 0.19 dB for Mobile, 0.66 dB for News, and 0.41 dB for Paris, with respect to those of the JVT-G012.

Fig. 7 illustrates the frame-to-frame quality (PSNR of luminance) comparison of the proposed rate-control scheme with the JVT-G012 rate-control scheme. Observe that much better improvements are achieved for Container, News, and Paris than that for Foreman and Mobile. The performance degradation is due to poor spatial-temporal correlations when picture activities are high.

The proposed rate-control scheme has also been tested using other target bit rates, frame rates, and other video sequences. The test result is summarized in JVT-O016 tests.doc, which can be referred to at http://ftp3.itu.ch/av-arch/jvt-site/2005_04_Busan. The subset of results presented in this paper is representative of where differences in performance were found among the different rate-control methods.

TABLE III
AVERAGE Y PSNR AND BIT-RATE COMPARISONS

Sequence	scheme		Initial Quantization Parameter										
			24	26	28	30	32	34	36	38	40	42	44
Container	JM 10	Y PSNR	38.37	36.92	35.56	34.38	33.1	31.7	30.27	29	27.76	26.52	
		Bit rate	16.29	11.39	8.17	5.92	4.41	3.35	2.54	1.93	1.53	1.2	
	Proposed	Y PSNR	38.83	37.43	36.03	34.79	33.38	32.02	30.61	29.27	27.83	26.62	
		Bit rate	16.27	11.39	8.16	5.91	4.41	3.35	2.55	1.93	1.53	1.2	
	Gain	Y PSNR	0.46	0.51	0.47	0.41	0.28	0.32	0.34	0.27	0.07	0.1	
Foreman	JM 10	Y PSNR	38.44	37.09	35.77	34.44	33.14	31.97	30.73	29.52	28.36	27.16	26.08
		Bit rate	42.57	32.37	24.75	18.88	14.35	11.19	8.59	6.71	5.33	4.15	3.31
	Proposed	Y PSNR	38.54	37.14	35.85	34.54	33.27	32.1	30.89	29.67	28.48	27.24	26.11
		Bit rate	42.59	32.38	24.75	18.84	14.32	11.16	8.58	6.69	5.31	4.14	3.29
	Gain	Y PSNR	0.1	0.05	0.08	0.1	0.13	0.13	0.16	0.15	0.12	0.08	0.03
Mobile	JM 10	Y PSNR	36.93	35.33	33.71	32.02	30.28	28.66	26.93	25.42			
		Bit rate	2866.87	2217.37	1673.65	1206.9	840.79	578.95	379.61	259.05			
	Proposed	Y PSNR	37.23	35.55	33.89	32.16	30.43	28.81	27.12	25.61			
		Bit rate	2871.66	2219.28	1674.36	1207.13	840.79	578.93	379.44	258.61			
	Gain	Y PSNR	0.3	0.22	0.18	0.14	0.15	0.15	0.19	0.19			
News	JM 10	Y PSNR	39.52	37.88	36.32	34.85	33.22	31.72	30.19	28.98	27.6	26.38	
		Bit rate	23.42	18.58	14.67	11.52	9.01	7.12	5.48	4.28	3.33	2.57	
	Proposed	Y PSNR	39.73	38.17	36.94	35.42	33.98	32.64	31.09	29.79	28.42	27.04	
		Bit rate	23.41	18.56	14.66	11.52	8.99	7.1	5.44	4.25	3.32	2.56	
	Gain	Y PSNR	0.21	0.29	0.62	0.57	0.76	0.92	0.9	0.81	0.82	0.66	
Paris	JM 10	Y PSNR	38.77	37.23	35.59	33.96	32.41	30.94	29.53	28.18	26.96		
		Bit rate	243.95	194.66	153.49	118.45	89.64	68.13	49.99	37.11	27.77		
	Proposed	Y PSNR	38.99	37.65	36.21	34.56	33.01	31.53	29.91	28.42	26.97		
		Bit rate	243.93	194.57	153.36	118.09	89.22	67.8	49.75	36.98	27.68		
	Gain	Y PSNR	0.22	0.42	0.62	0.6	0.6	0.59	0.38	0.24	0.01		

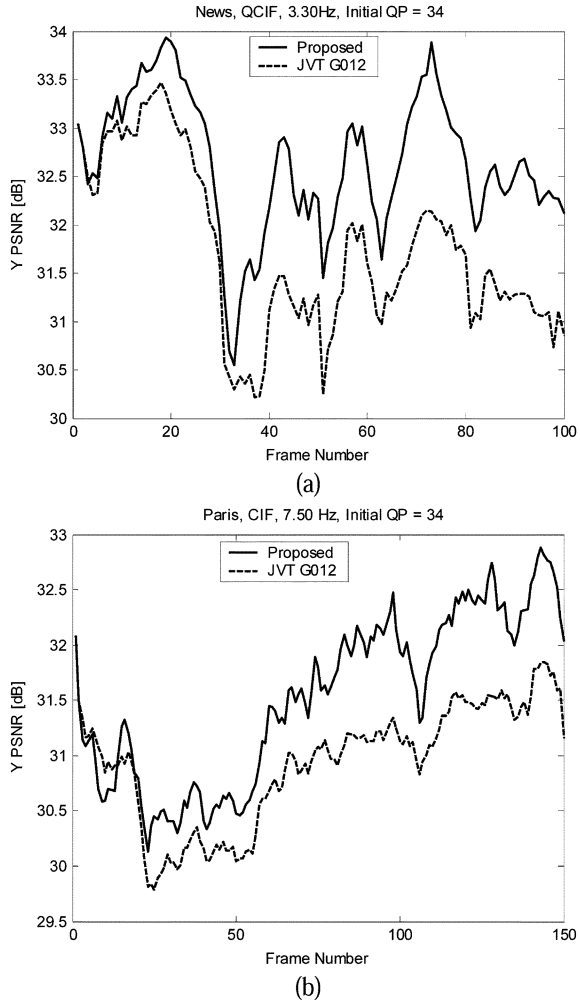


Fig. 7. PSNR per frame for the proposed scheme versus JVT-G012.

VII. CONCLUDING REMARKS

In this work, a novel adaptive coding characteristics prediction scheme is presented to improve the accuracy of R-D modeling, by exploiting spatio-temporal correlations. With the proposed prediction scheme, we present a modified Q2–rate function and a linear distortion model, and then deduce a simple closed-form solution to the problem of optimum bit allocation, just in a TMN-8-alike way. The experimental results show that it is more accurate and robust than JVT-G012, the current standardized rate-control scheme.

Nevertheless, it is worthy to point out that the proposed prediction scheme maybe suffers from performance degradation at scene changes due to the assumption of spatial-temporal correlations, because this assumption no longer holds at this time.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their useful comments and suggestions. They would also like to thank Y. Shen for his useful suggestions.

REFERENCES

[1] *Generic Coding of Moving Pictures and Associated Audio Information—Part 2: Video*, ITU-T and ISO/IEC JTC1, ITU-T Recommend. H.262 - ISO/IEC 13818-2 (MPEG-2), Nov. 1994.

[2] *Video Coding for Low Bitrate Communication*, ITU-T, ITU-T Recommend. H.263; version 1, Nov. 1995, version 2, Jan. 1998; version 3, Nov. 2000.

[3] *Coding of Audio-Visual Objects—Part 2: Visual*, ISO/IEC JTC1, ISO/IEC 14496-2 (MPEG-4 visual version 1), Apr. 1999; Amendment 1 (version 2), Feb. 2000; Amendment 4 (streaming profile), Jan. 2001.

[4] T. Wiegand and G. J. Sullivan, "Draft ITU-T recommendation H.264 and final draft international standard of joint video specification (ITU-T recommendation H.264|ISO/IEC 14496-10 AVC)," Joint Video Team of ISO/IEC JTC1/SC29/WG11 and ITU-T SG16/Q.6 Doc. JVT-G050 Thailand, Pattaya, Mar. 2003.

[5] 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.

[6] J. Ribas-Corbera and S. Lei, "Rate control in DCT video coding for low-delay communications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 1, pp. 172–185, Feb. 1999.

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

[8] 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.

[9] L.-J. Lin and A. Ortega, "Bit-rate control using piecewise approximated rate-distortion characteristics," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 38, no. 4, pp. 82–93, Jan. 1990.

[10] 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. 1, pp. 246–250, Feb. 1997.

[11] *Video Codec Test Model*, ITU-T/SG15, TMN8, Portland, OR, Jun. 1997.

[12] Test Model Editing Committee, MPEG-2, Test Model 5, Doc. ISO/IEC JTC1/SC29 WG11/93-225b., Apr. 1993.

[13] Coding of Moving Pictures and Associated Audio MPEG 97/W1796, Text of ISO/IEC 14496-2 MPEG-4 Video VM-Version 8.0, ISO/IEC JTC1/SC29/WG11, Video Group, Stockholm, Sweden, Jul. 1997.

[14] T. Wiegand, H. Schwarz, A. Joch, and G. Sullivan, "Rate constrained coder control and comparison of video coding standards," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 688–703, Jul. 2003.

[15] S. W. Ma, W. Gao, Y. Lu, and H. Q. Lu, Proposed draft description of rate control on JVT standard Joint Video Team of ISO/IEC JTC1/SC29/WG11 and ITU-T SG16/Q.6 Doc. JVT-F086, Awaji, Japan, Dec. 2002.

[16] Z. G. Li, F. Pan, K. P. Lim, G. N. Feng, X. Lin, and S. Rahardaj, Adaptive basic unit layer rate control for JVT, Joint Video Team of ISO/IEC JTC1/SC29/WG11 and ITU-T SG16/Q.6 Doc. JVT-G012, Pattaya, Thailand, Mar. 2003.

[17] J. Chalidabhongse and C.-C. Jay Kuo, "Fast motion vector estimation using multiresolution-spatio-temporal correlations," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 3, pp. 477–488, Jun. 1997.

[18] S. Zafar, Y.-Q. Zhang, and J. S. Baras, "Predictive block-matching motion estimation for TV coding—Part I: Inter-block prediction," *IEEE Trans. Broadcast.*, vol. 37, no. 3, pp. 97–101, Sep. 1991.

[19] S. Zafar, Y. Q. Zhang, and B. Jabbari, "Multiscale video representation using multiresolution motion compensation and wavelet decomposition," *IEEE J. Sel. Areas Commun.*, vol. 11, no. 1, pp. 24–35, Jan. 1993.

[20] Y.-Q. Zhang and S. Zafar, "Predictive block-matching motion estimation for TV coding—Part II: Inter-frame prediction," *IEEE Trans. Broadcast.*, vol. 37, no. 3, pp. 102–105, Sep. 1991.

[21] H. J. Lee, T. Chiang, and Y. Q. Zhang, "Scalable rate control for MPEG-4 video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 6, pp. 878–894, Sep. 2000.

[22] A. Viterbi and J. Omura, *Principles of Digital Communication and Coding*. New York: McGraw-Hill Electrical Engineering Series, 1979.

[23] D. A. Pierre, *Optimization Theory with Applications*. New York: Dover, 1986.

[24] JM 10 [Online]. Available: http://iphome.hhi.de/suehring/tml/download/old_jm/jm10.zip

[25] Z. He and S. K. Mitra, "Optimum bit allocation and accurate rate control for video coding via ρ domain source modeling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 10, pp. 840–849, Oct. 2002.

[26] G. J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE Signal Process. Mag.*, vol. 15, no. 6, pp. 74–90, Nov. 1998.

[27] B. Tao, B. W. Dichinson, and H. A. Peterson, "Adaptive model-driven bit allocation for MPEG video coding [J]," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 1, pp. 147–157, Feb. 2000.

Wu Yuan received the B.S. degree in computer science from Southeast University, Nanjing, China, in 1997, and the M.S. degree in 2000 from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, where he is currently pursuing the Ph.D. degree.

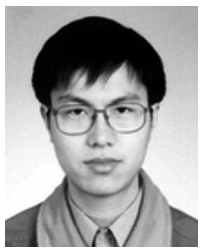
He was with Bell Labs (China) and Lucent Technologies, Beijing, as a Member of Technical Staff from 2000 to 2001. He was with CASW Data Technology, Beijing, as a Software Engineer from 2001 to 2002. He was with PCS, Motorola (China), Beijing, from 2002 to 2003. His research interests include data networking, data warehouse, grid computing, and video coding.

Shouxun Lin (M'99) received the Ph.D. degree from Beijing University of Technology, Beiging, China, in 1998.

Since 1995, he has been Associate Professor and Professor (in 2000) with the Institute of Computing Technology, Chinese Academy of Sciences (CAS), Beijing. From 2000 to 2005, he was Vice Director of the Digital Laboratory at the Institute of Computing Technology, CAS. His research interests include multimedia processing and comparison, video coding, video analysis, multimedia indexing, statistical machine translation, and evaluation of computer human interaction.

Yongdong Zhang received the Ph.D. degree in electronic engineering from Tianjing University, Tianjing, China, in 2002.

He is an Associate Professor with the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China. His research interests are in the field of video coding and transcoding, video analysis and retrieval, and universal media access.



Wen Yuan received the Ph.D. degree in geographical information science from Peking University, Peking, China, in 2004.

Currently, he is a Postdoctoral Researcher with the Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences. His research interests include grid computing, 3-D, global grid model, spatial analysis, and spatial database.

Haiyong Luo received the B.S. degree in information engineering from Huazhong University, Wuhan, China, in 1989 and the M.S. degree from Beijing University of Posts and Telecommunications, Beijing, China, in 2002, and is currently pursuing the Ph.D. degree at the Institute of Computing Technology, Chinese Academy of Sciences, Beijing.

His research interests include video coding, embedded systems, data networking, and wireless communication.