Rate-Distortion Analysis of Dead-Zone Plus Uniform Threshold Scalar Quantization and Its Application—Part II: Two-Pass VBR Coding for H.264/AVC

Jun Sun, *Member, IEEE*, Yizhou Duan, Jiangtao Li, Jiaying Liu, *Member, IEEE*, and Zongming Guo, *Member, IEEE*

***Abstract*— In the first part of this paper, we derive a source model describing the relationship between the rate, distortion, and quantization steps of the dead-zone plus uniform threshold scalar quantizers with nearly uniform reconstruction quantizers for generalized Gaussian distribution. This source model consists of rate-quantization, distortion-quantization (D-Q), and distortion-rate (D-R) models. In this part, we first rigorously confirm the accuracy of the proposed source model by comparing the calculated results with the coding data of JM 16.0. Efficient parameter estimation strategies are then developed to better employ this source model in our two-pass rate control method for H.264 variable bit rate coding. Based on our D-Q and D-R models, the proposed method is of high stability, low complexity and is easy to implement. Extensive experiments demonstrate that the proposed method achieves: 1) average peak signal-to-noise ratio variance of only 0.0658 dB, compared to 1.8758 dB of JM 16.0's method, with an average rate control error of 1.95% and 2) significant improvement in smoothing the video quality compared with the latest two-pass rate control method.**

***Index Terms*— H.264/AVC, rate-distortion theory, two-pass rate control, video coding.**

I. Introduction

An IMPORTANT contribution of rate-distortion (R-D) analysis is the source model describing the relationship between bit rate, distortion and quantization step. Besides the theoretical validation in simulation experiments, the effectiveness of the source model is further required to be verified in practical video applications. In practice, video encoders produce two types of applications: constant bit rate (CBR) and variable bit rate (VBR). To well support both CBR and VBR video applications, rate control methods play an essential role. In CBR coding, single-pass rate control is the only feasible method in the bandwidth and complexity constrained situations to help match the target rate. However, the information collected within the single-pass coding is not enough to well reflect the entire sequence, which inevitably leads to frequent fluctuation of video quality. Therefore, for most off-line applications where video quality is mainly concerned and the constraint of instantaneous rate or coding delay is not that strict, VBR coding is a better choice since more sophisticated two-pass rate control strategies can be employed to achieve higher coding efficiency and smoother visual quality under the target average rate. Our work focuses on the effective application of the source model in two-pass rate control for H.264/AVC [1].

Normally, the entire two-pass rate control method is performed in two steps: 1) characteristic statistics of the whole video sequence and 2) bit allocation for final coding. The first step is finished within the first-pass encoding, where necessary statistic data of all frames are obtained. Then in the second step, the collected information of the whole sequence is used to initialize the source model, by which corresponding strategies are established to determine the quantization parameter (QP) for each frame before or within the second-pass encoding, so that the desired average rate and visual quality can be achieved. In VBR coding, the key indexes to evaluate the performance of a two-pass rate control method are the smoothness of reconstructed video quality and the rate control accuracy. More specifically, using PSNR as the distortion measure, a two-pass rate control method can be regarded effective if for any given video sequence, all its coded frames are of similar PSNR while the average rate control error is within an acceptable range. In addition, the coding efficiency denoted by average PSNR of the coded sequence under the method is also observed as a secondary evaluation criterion.

In general, all the two-pass rate control methods can be classified into two categories. In the first category, the bit allocation of each frame is performed during the second-pass encoding, so that the encoder can flexibly monitor and minimize the rate control error. Lie *et al.* [2] used a window-based method to segment the whole sequence into several windows and built up the R-D model with Lagrange multiplier to coordinate the QP of every macro block in the second-pass encoding. Similar method was applied by Kwon *et al.* [3] with

Manuscript received October 12, 2011; revised May 10, 2012 and August 13, 2012; accepted August 20, 2012. Date of publication August 31, 2012; date of current version December 20, 2012. This work was supported by the National Natural Science Foundation of China under Contract 60902004, the National Basic Research Program of China under Contract 2009CB320907, and the Ministry of Education of China through the Doctoral Fund under Contract 20110001120117. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Ali Bilgin.

The authors are with the Institute of Computer Science and Technology, Peking University, Beijing 100871, China (e-mail: liujiaying@pku.edu.cn).

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

Digital Object Identifier 10.1109/TIP.2012.2216275

the enhancement of introducing a GOP-level adjust scheme to improve the rate control performance. Zhang *et al.* [4] proposed a hybrid two-pass rate control method for H.264 high definition video coding based on their work in [5]. By together employing scene change detection with the GOP-level complexity model and the linear distortion-quantization (D-Q) model, bit rate and video quality can be adjusted adaptively according to the actual status of the second-pass encoding.

Unlike the first category two-pass rate control method, in the second category, the bit allocation of each frame is performed during the off-line processing, and then in the second-pass encoding each frame is simply coded with the predetermined QP without adjustment. In this case, the rate control performance is totally depending on the source model, which actually decides the QP values for final encoding. In [6]–[8], several rate-quantization (R-Q) models are proposed to calculate the QP values. These R-Q models [6]–[8] are respectively obtained via visual perception experiments under the given resolution and frame rate, through weighted adjustment for I-, P- and B-frames base on TM5 test model [9], or from the first-pass CBR encoding statistics. The implementation of the distortion-rate (D-R) and D-Q models in two-pass rate control is represented by Huang *et al.*'s method [10], where the D-R model calculates the corresponding distortion under the target rate, and the D-Q model derives the QP values accordingly.

As is introduced above, all the previous work has provided useful references for two-pass rate control video applications. However, these contributions are still insufficient to satisfy the requirement of effective VBR coding. For the first category methods, there is no available solution to avoid or minimize the fluctuation of video quality [4]. In contrast, the second category methods provide smoother visual quality but result in more obvious rate control error [4]. The most important reason for the drawbacks of the two categories is that the source models applied by the previous work fail to well represent the characteristics of the real video signals, which seriously affects the subsequent bit allocation strategies. The high computational complexity of the empirical source models represented by Westrink *et al.*'s method [6] makes them less practical and appealing. And the lack of rigorous validation of the analytical and heuristic source models in video applications leads to inaccurate estimated rate and distortion, which explains why rate control accuracy and constant video quality cannot be simultaneously guaranteed by the existing rate control methods.

In this paper, we propose a simple but efficient two-pass rate control method for H.264 VBR coding completely based on the D-R and D-Q models developed in Part I of this paper. Before introducing our rate control method, we first compare the numerical results of our R-Q, D-Q and D-R models with the actual coding data produced by H.264/AVC reference software JM 16.0 [11] for two purposes: 1) to confirm the effectiveness of the proposed models and 2) to inspire the strategy to estimate the parameters of these models in application. We explain in detail why the R-Q model cannot be directly used to efficiently predict the actual coding rate in practice, and give suggestions to the parameter estimation of the D-R and D-Q models to make sure they can effectively benefit the rate control method. The proposed method belongs to the second category mentioned above, where both of the D-R and D-Q models are utilized to optimize the bit allocation in the off-line processing. The effectiveness of all our findings is finally justified by the experimental results of multiple standard test sequences covering various resolutions and target rates.

The rest of this paper is organized as follows. The proposed source model of the dead-zone plus uniform threshold scalar quantizers with nearly uniform reconstruction quantizers (DZ+UTSQ/NURQ) [12] for generalized Gaussian distribution (GGD) [13] is briefly reviewed in Section II. Then in Section III, the effectiveness of the proposed R-Q and D-Q models are further verified in H.264/AVC, and the parameter estimation strategies for the effective application of the proposed D-Q and D-R models are provided in Section IV. On this basis, the novel two-pass rate control method for H.264/AVC is developed in Section V, followed by the experimental results and discussion in Section VI. Finally, we give conclusions in Section VII.

II. Source Model of DZ + UTSQ/NURQ for GGD

In Part I of this paper, by using analytical and heuristic methods, we derive a source model of DZ+UTSQ/NURQ for GGD, describing the relationship between rate, distortion and quantization step. The source model consists of R-Q, D-Q and D-R models. The proposed R-Q model is expressed as:

$$H(\Delta) = -\frac{\ln(1-e^{-z(\sqrt{2}\Delta/\beta)^{\alpha}})}{\ln 2} + e^{-z(\sqrt{2}\Delta/\beta)^{\alpha}} + \frac{(\sqrt{2}\Delta/\beta)^{\alpha}\cdot e^{-z(\sqrt{2}\Delta/\beta)^{\alpha}}\left(z + e^{-(\sqrt{2}\Delta/\beta)^{\alpha}} - z e^{-(\sqrt{2}\Delta/\beta)^{\alpha}}\right)}{\left[(1-e^{-(\sqrt{2}\Delta/\beta)^{\alpha}})\cdot \ln 2\right]}. \tag{1}$$

$H(\Delta)$ represents the output entropy rate of encoder, measured by bits/sample. And the proposed D-Q model is:

$$D(\Delta) = \beta^2\Bigg[1 + \frac{e^{-z\left(\frac{\sqrt{2}\Delta}{\beta}\right)^{\alpha}}}{2\left(1-e^{-\left(\frac{\sqrt{2}\Delta}{\beta}\right)^{\alpha}}\right)} \times \left[\left(\frac{\sqrt{2}\Delta}{\beta}\right)^{2\alpha}(1-2z) - 2\left(\frac{\sqrt{2}\Delta}{\beta}\right)^{\alpha}\right]\Bigg] \tag{2}$$

which describes the distortion originating from quantization process under the mean square error (MSE) criterion. The proposed D-R model was first developed in [14] for MPEG-4 FGS video coding [15] and extended to the general form of DZ+UTSQ/NURQ as a more universal conclusion:

$$PSNR(R) = aR + A - (A-B)/(1+bR). \tag{3}$$

For the R-Q and D-Q models, α is the GGD shape parameter, β is the standard deviation of GGD, Δ is quantization step and z is the dead-zone ratio of DZ+UTSQ/NURQ. For the

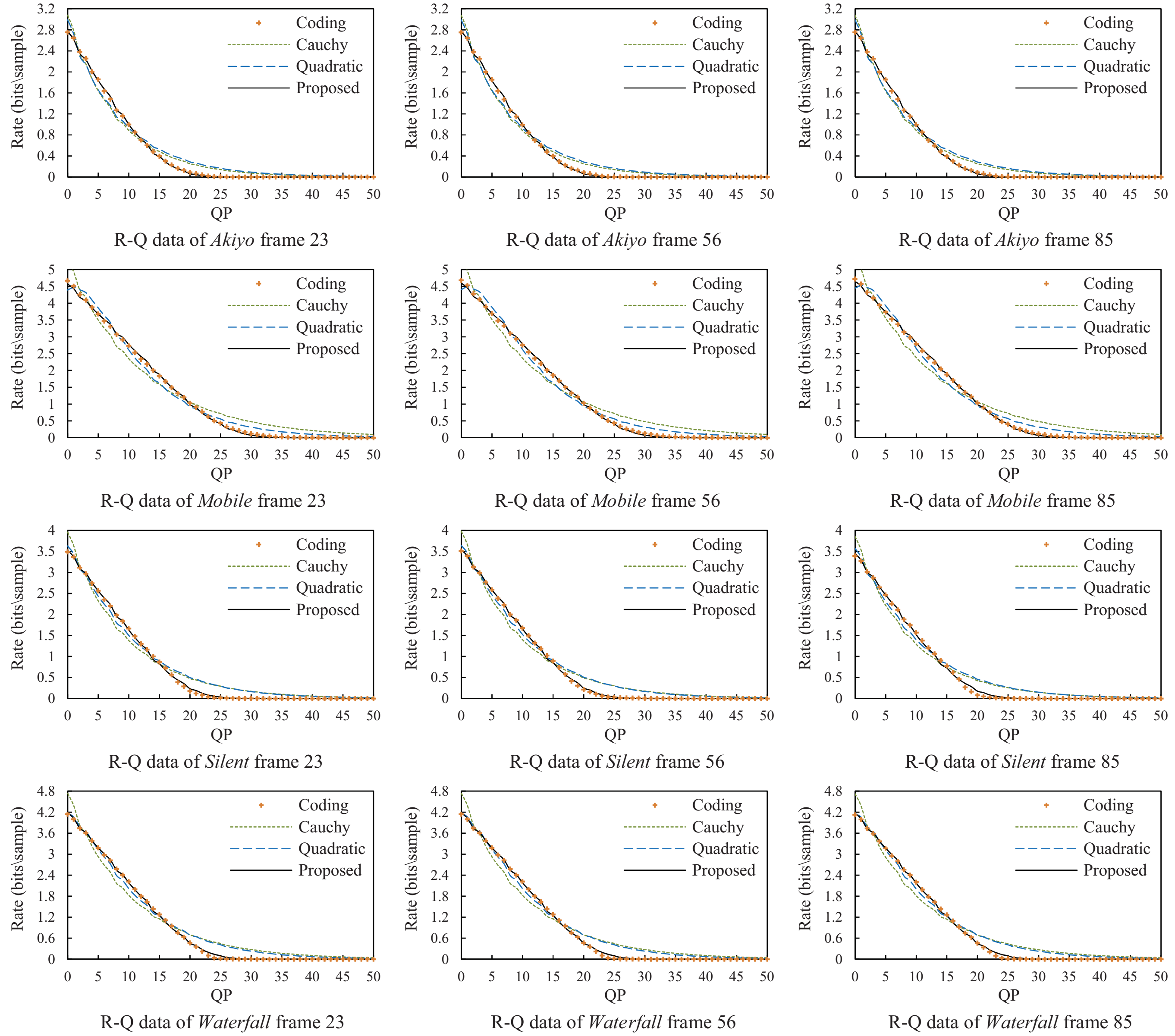


Fig. 1. R-Q model validation by comparing the actual coding rate of JM 16.0 with the estimated rate of the three source models.

D-R model, R is the entropy rate, B is the distortion measured in PSNR when $R = 0$, a and A are the asymptote parameters, and b controls the approach of the actual D-R function to the asymptote.

III. Evaluation of the Proposed R-Q and D-Q Models in H.264/AVC

As is shown in Part I of this paper, the proposed GGD R-Q and D-Q models are of superior accuracy in estimating the theoretical rate and distortion. In this section, by further comparing the numerical results of different models based on the actual coding data produced by JM 16.0, the effectiveness of the proposed R-Q and D-Q models is rigorously confirmed.

First and foremost, we explain the design of the validation experiment by proclaiming the prerequisite of the source model. The prerequisite of the source model is the stationary source. In R-D modeling, the source model describes the relationship between rate, distortion and quantization step for stationary input signals. Therefore in practical video coding systems, the actual input signals to the quantizer, known as the residue transform coefficients, are required to keep unchanged when the output rate and distortion are observed under different QP values. In order to approximate this prerequisite, in the validation experiments we first choose a target frame and change its encoding QP to obtain different rate and distortion. Meanwhile, we fix most of the other data dependencies of the target frame, such as reference frame number, macro block (MB) type and the QP used for R-D optimization (RDO) process, so that the effect of different encoding QP on the residue transform coefficients of the target frame is minimized. However, it should be noted that for the target frame containing intra MBs, different QP values still have significant influence on the residue transform coefficients. In intra prediction, the

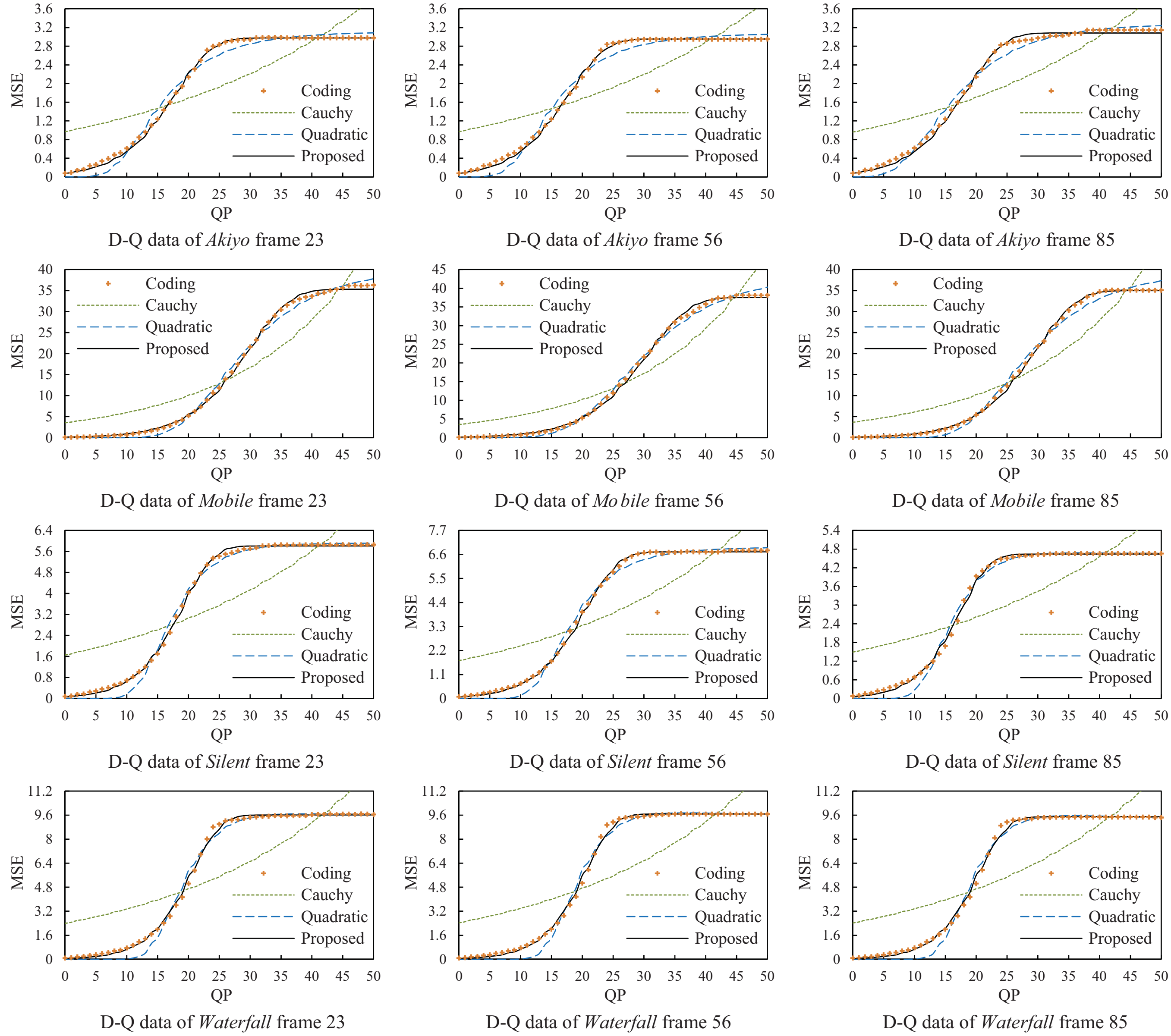


Fig. 2. D-Q model validation by comparing the actual coding rate of JM 16.0 with the estimated rate of the three D-Q models.

reconstructed pixels of neighboring MBs are used as predictors which are of totally different distortion under different QP, thus the prerequisite of the source model actually cannot be guaranteed. In Section IV, the error caused by intra MB in estimating the parameters of the proposed R-Q and D-Q models is investigated in detail.

In the validation experiments, four standard test sequences *Akiyo*, *Mobile*, *Silent* and *Waterfall* (CIF format, 30 frame/s) are encoded, covering various motion types. For each sequence, we randomly select 10 target frames to observe. For comparison, we still choose the Cauchy model [16] and the Quadratic model [17] to keep consistent with the theoretical validation in Part I of this paper. And the non-linear data fitting method is employed to obtain the parameters of each model based on the actual data produced by JM 16.0 under QP value from 0 to 50. Important configurations of JM 16.0 encoder are listed as follows: High Profile is used; Adaptive Rounding is turned off so that the default setting of the dead-zone ratio z in H.264/AVC is applied (2/3 for Intra frames and 5/6 for Inter frames [18]); UMH exagon fast motion search algorithm is used and the search range is 65 by 65; the number of reference frame is 2 and the frame structure is IBBP; the entropy coding method is the context adaptive binary arithmetic coding (CABAC); the high complexity RDO mode is applied and all available modes for both Intra and Inter frames are enabled. The main reason we turn off Adaptive Rounding is to apply the same setting of z in the proposed source model and JM 16.0, so that the accuracy of the source model can be rigorously verified. Since Adaptive Rounding only effectively improves the coding efficiency in very high bit rate [19], this configuration actually does not affect the correctness of the validation experiments.

To comprehensively reflect the effectiveness of the proposed source model, the validation results are exhibited from two

TABLE I
COMPARISON OF FITTING ACCURACY OF THE THREE R-Q MODELS

Sequence	Measures	Goodness of R-Q Models		
		Proposed	Cauchy	Quadratic
Akiyo	RMSE	0.0158	0.1139	0.1195
	R-square	0.9996	0.9801	0.9781
Mobile	RMSE	0.0367	0.2725	0.1491
	R-square	0.9994	0.9679	0.9903
Silent	RMSE	0.0301	0.1962	0.1636
	R-square	0.9992	0.9672	0.9769
Waterfall	RMSE	0.0309	0.2604	0.1811
	R-square	0.9995	0.9620	0.9816
Average	RMSE	0.0284	0.2108	0.1533
	R-square	0.9994	0.9693	0.9817

TABLE II
COMPARISON OF FITTING ACCURACY OF THE THREE D-Q MODELS

Sequence	Measures	Goodness of R-Q Models		
		Proposed	Cauchy	Quadratic
Akiyo	RMSE	0.0476	0.6346	0.1403
	R-square	0.9982	0.6951	0.9853
Mobile	RMSE	0.4845	5.5263	0.9584
	R-square	0.9990	0.8641	0.9959
Silent	RMSE	0.0800	1.3083	0.2239
	R-square	0.9987	0.6816	0.9909
Waterfall	RMSE	0.1758	2.2917	0.4199
	R-square	0.9981	0.6884	0.9897
Average	RMSE	0.1970	2.4402	0.4356
	R-square	0.9985	0.7323	0.9905

perspectives. In Fig. 1 and Fig. 2, the numerical results of the three R-Q and D-Q models in target frames 23, 56 and 85 of the four sequences are intuitively illustrated. It should be noted that the entropy rate is measured in bits/sample, which means even slight difference can result in significant prediction error in the bit rate of the whole system. Therefore, it is clearly exhibited that the proposed R-Q and D-Q model matches the rate and distortion of the practical video coding systems perfectly from low to high bit rate while other two models are both of some deviations. In Table I and Table II, the goodness of the three R-Q and D-Q models is evaluated under two representative measures: root of MSE (RMSE) [20] and coefficients of determination (R-Square) [21]. And the result closer to 0 under RMSE or closer to 1 under R-Square indicates a better fitting. The fitting evaluation covers all the target frames in the four test sequences, being a convincing support to show the superior accuracy of the proposed R-Q and D-Q models in estimating the rate and distortion of the practical video coding systems.

It should also be noted that in Table II, the Cauchy D-Q model is observed to deviate more obviously from the actual distortion than the quadratic model in the fitting evaluation, which is mainly due to the following two reasons. First, as we have proclaimed, the validation experiments are designed to evaluate the source models under the prerequisite of stationary source. In this case, the maximum distortion of the target frame equals to the variance of the residue transform coefficients, which can be achieved in very low bit rate (large QP). In contrast, the Cauchy model is proposed to well describe the source without variance. As a result, it may fail to fit the actual data with variance from low to high bit rate. In practical video coding, however, it is more likely to observe the distortion within the suitable bit rate range for application, which is also a more suitable environment to employ the Cauchy model. Second, compared with the Cauchy model, the quadratic D-Q model has one more parameters to further refine the model calculation, thus providing decent data fitting goodness.

IV. SOURCE MODEL PARAMETER ESTIMATION

With the validation experiments over the simulated theoretical data in Part I of this paper and the actual coding data in Section III, the effectiveness of the proposed R-Q and D-Q models is rigorously verified. However, for the implementation of the proposed R-Q and D-Q models in practical video applications, it is in capable to obtain the optimal GGD parameters α and β using the non-linear data fitting method. Even in two-pass VBR coding, it is a typical ill-posed inverse problem to obtain the optimal GGD parameters only with the first-pass information. In this section, we first introduce the basic GGD parameter estimation method and evaluate its performance by comparing with the optimal parameters derived from data fitting. Based on the motivating observations, the efficient model parameter estimation strategies are developed for the proposed D-Q and D-R models, providing guidelines for the effective application of the source model in two-pass rate control method for H.264/AVC.

A. Basic GGD Parameter Estimation Method

The basic GGD parameter estimation (BGPE) method is first proposed in [22] and is further simplified in [23] as follows: with the random variable X as source signals, the parameters of GGD can be estimated by

$$\alpha^* = F\left(\frac{\left(\frac{1}{N}\sum_{k=1}^{N}|X_k|\right)^2}{\frac{1}{N}\sum_{k=1}^{N}X_k^2}\right) \quad \text{and} \quad \beta^* = \sqrt{\frac{1}{N}\sum_{k=1}^{N}X_k^2} \tag{4}$$

where, $F(x) = 0.2718/(0.7697 - x) - 0.1247$, α^* and β^* are the estimated values of α and β. In the hybrid transform-based video coding framework, the source signals X denote the residue transform coefficients.

It should be noted that although the BGPE method has been indicated to well simulate the distribution of the given residue transform coefficients in H.264 [10], [14], [24] its effectiveness is entirely based on the stable source signals. In two-pass

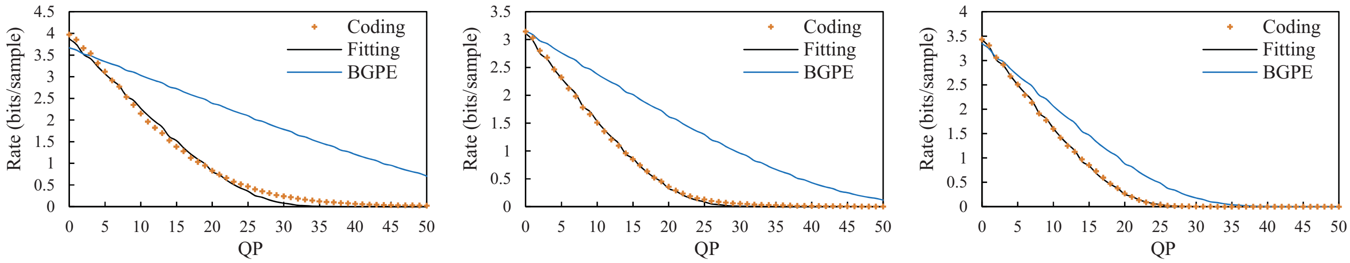


Fig. 3. Evaluation of employing the BGPE method in the proposed R-Q model compared with the optimal fitting rate and the coding rate of JM 16.0.

TABLE III
COMPARISON OF GGD PARAMETERS DERIVED FROM DATA FITTING AND BGPE METHOD FOR THE PROPOSED R-Q MODEL

Frame Number	Frame Type	Data Fitting		BGPE Method	
		α	β	α	β
0	I	0.800	2.429	0.316	7.819
78	P	0.796	1.786	0.371	4.224
38	B	0.914	1.991	0.638	2.622

VBR coding,since the first-pass and second-pass encodings use the same video sequence as input, it is natural to directly use the first-pass source information to estimate α and β and determine the second-pass encoding QP, with the assumption that the source statistics in the first and second-pass encoding remain the same. However, according to the practical coding requirement, it is very likely that the determined second-pass encoding QP is quite different from the first-pass, in which case the actual source statistics in the second-pass encoding may change drastically. In the following subsections, the effectiveness of employing the BGPE method in the proposed R-Q and D-Q models under various QP is carefully evaluated. And efficient parameter estimation strategies are developed for the proposed D-Q and D-R models in two-pass VBR coding.

B. GGD Parameter Estimation for the Proposed R-Q Model

To well reflect the performance of the BGPE method over the actual video source signals, the target frames with frame type I-, P- and B- are observed, each containing different numbers of intra MBs. We employ the experimental settings described in Section III-A to obtain the coding rate of JM 16.0. Then we use the proposed R-Q model to produce the calculated rate with two groups of GGD parameters: the optimal parameters and the estimated parameters. The optimal parameters are derived by using the non-linear data fitting method base on the coding rate, and the estimated parameters are counted by averaging the BGPE results under each QP value. Because of the limited space, we just illustrate the comparison results of the coding rate and the two calculated rates over *Foreman* sequence (CIF format, 30 frame/s) in Fig. 3. It is clear that with the α and β derived using the BGPE method, the calculated rate of the R-Q model has obvious deviation from the coding rate of JM 16.0 in B-frame, and this mismatch is amplified in P-frame and maximized in I-frame. In Table III, for each frame, the GGD parameters derived from data fitting and BGPE method are listed respectively, from which significant variance can be seen in both α and β. With above observations, the BGPE method is concluded not to efficiently approximate the optimal parameters for the proposed R-Q model.

In actual video coding, the BGPE method of (4) is not fit for the proposed R-Q model mainly due to the following two reasons. First and foremost, the entropy coding method applied in encoders, either the CABAC or the context adaptive variable length coding (CAVLC), is somewhat inferior to the theoretical entropy compression efficiency. Therefore, the entropy coded bits, namely the coding rate, actually do not match the calculated rate of the R-Q model with the estimated parameters. Second, as is mentioned in Section IV-A, when QP changes, the statistics of the actual video source signals are different. To be specific, for intra MBs, with the increment of QP, the growing reconstruction distortion in the neighboring MB results in totally different residue transform coefficients of the current MB through the error propagation in intra prediction. The estimated parameters from the BGPE method present different sources under different QP, while the optimal parameters provide a most suitable source to reflect the overall coding rate. This also well explains why more obvious variance of the estimated and the optimal parameters is observed in the frames containing more intra MBs (such as I- and P-frames).

For the H.264/AVC two-pass VBR coding, all the frames are encoded merely once in the first-pass encoding, thus only one output entropy rate and QP can be collected for each frame, with which the data fitting method is not able to be applied. Meanwhile, the BGPE method is not a good approximation of the data fitting method in calculating the coding rate. As a result, the proposed R-Q model is not utilized in our two-pass rate control design for H.264/AVC.

C. GGD Parameter Estimation for the Proposed D-Q Model

Similarly, we observe the performance of the BGPE method in the proposed D-Q model under various QP values by comparing the estimated and optimal GGD parameters via the experiments mentioned in Section IV-B. The coding distortion of JM 16.0 and the calculated distortion from the proposed D-Q model are measured under the MSE criterion. It is noted that for P-frame, the D-Q data is only observed within the QP range of 0–44, while forI-frames this range is further narrowed

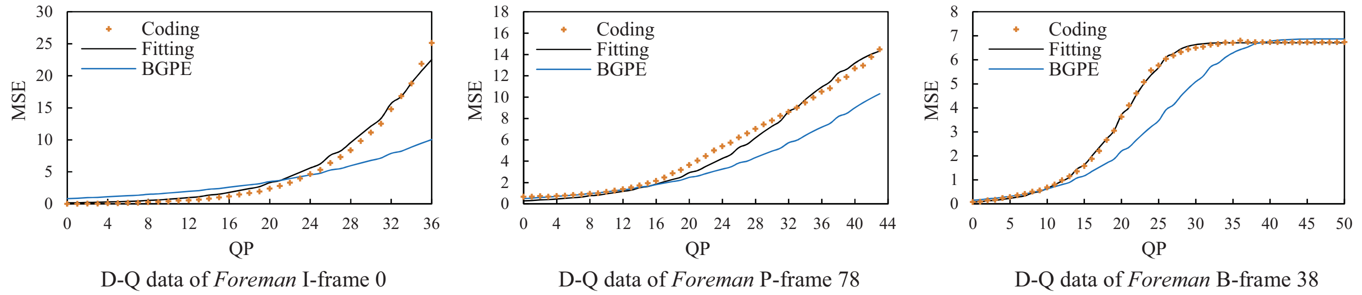


Fig. 4. Evaluation of employing the BGPE method in the proposed D-Q model compared with the optimal fitting distortion and the coding distortion of JM 16.0.

TABLE IV
COMPARISON OF GGD PARAMETERS DERIVED FROM DATA FITTING AND BGPE METHOD FOR THE PROPOSED D-Q MODEL

Frame Number	Frame Type	Data Fitting		BGPE Method	
		α	β	α	β
0	I	0.616	5.399	0.316	7.819
78	P	0.542	3.976	0.371	4.224
38	B	0.959	2.590	0.638	2.622

to 0–36. The purpose we choose different QP ranges for different frame types is to control the reconstruction distortion in the frames containing large numbers of intra MBs, so that the change in the source statistics will not bring too much error to the GGD parameter estimation for the proposed D-Q model.

Related results of the comparison experiments are exhibited in Fig. 4, and the corresponding values of α and β derived from data fitting and BGPE method are listed respectively in Table IV. It is shown that when using α and β derived from the BGPE method, the calculated distortion still deviates from the actual coding distortion. However, from Table IV it is clear that for each frame, the difference between the optimal and estimated parameters is not as significant as it is in Table III, which is mainly because no extra mismatch is introduced in the comparison of the coding and calculated distortion. The D-Q model calculation denotes the distortion from quantization, while the reconstructed frame of JM 16.0 involves all the coding distortions which can also be approximated as the distortion from quantization [24]. Therefore, compared with the result of R-Q model, the BGPE method for the proposed D-Q model leads to the estimated GGD parameters closer to the optimal parameters derived from data fitting. According to Table IV, in particular for P- and B-frames, the data fitting method and the BGPE method obtain very similar β values, and for I-frame the deviation of the estimated β from the optimal value is also reduced.

Inspired by above observation and analysis, efficient GGD parameter estimation strategy is concluded to effectively implement the proposed D-Q model in the two-pass rate control design for H.264 VBR coding. In the first-pass encoding, the QP and MSE of each frame are collected and the BGPE method is applied only to obtain β. With the QP, MSE and β, the corresponding value of α is calculated by the proposed D-Q model. Using this strategy, the estimated

TABLE V
EMPIRICAL VALAUES OF TWO D-R MODEL PARAMETERS IN DIFFERENT TYPES OF FRAMES

Frame Type	*a*	*b*
I	5.0	10.5
P	2.5	10.0
B	4.5	4.8

α and β can well approximate the optimal GGD parameters derived from data fitting, so that the bit allocation is optimized in the off-line processing. It is also noted that the deviation of β in I-frame will not cause much model error, since the number of I-frames in practical video applications is restricted considering the coding efficiency.

D. Parameter Estimation for the Proposed D-R Model

Unlike the GGD R-Q and D-Q models which are deduced from the R-Q and D-Q functions of DZ+UTSQ/NURQ for GGD source, the proposed D-R model is developed mainly base on the analysis of GGD D-R derivative function, indicating the direct relationship between rate and distortion. By together collecting the coding rate and distortion of JM 16.0 in Section IV-B and IV-C, the actual D-R data of the selected frames is obtained. And we compare the actual D-R date with the calculated D-R data for each frame to develop the efficient parameter estimation strategies for the proposed D-R model.

To implement the proposed D-R model, the four model parameters introduced in Section II are required to be determined. Among these parameters, B represents the distortion under PSNR criterion when rate equals to zero. It is known that under MSE criterion, without the entropy rate from quantizer, the distortion equals to the variance of all the residue transform coefficients, denoted by β^2 for GGD sources. Thus with β obtained by the BGPE method, B is easily computed. For the other three model parameters a, b and A, two different simulation methods are employed: 1) a and b are both fixed to the empirical values given by Table V while A is obtained by the data fitting method, and 2) all of the three parameters are learned from data fitting. The actual D-R data of JM 16.0 and the calculated D-R data from the proposed D-R model are exhibited in Fig. 5 for I-, P- and B-frames, from which it is well illustrated that the proposed D-R model is a very precise approximation of the D-R relationship in real video coding. It is also observed that even when a and b are both fixed to

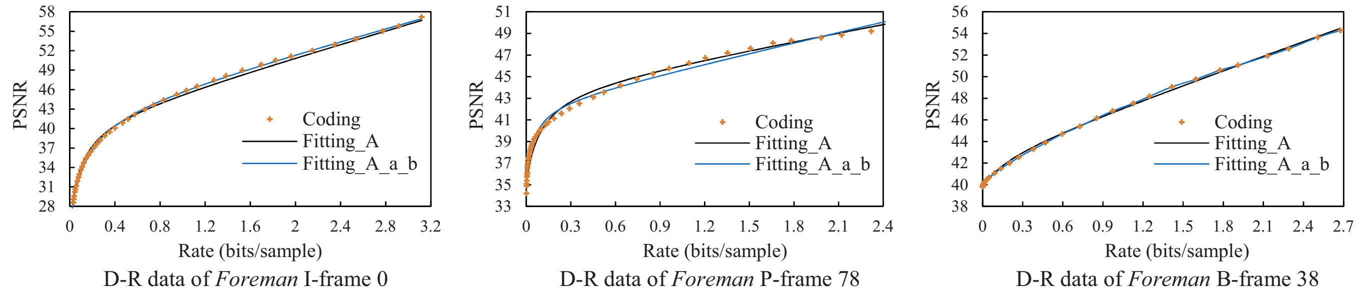


Fig. 5. Comparison of actual D-R data of JM 16.0 with calculated D-R data from the proposed D-R model.

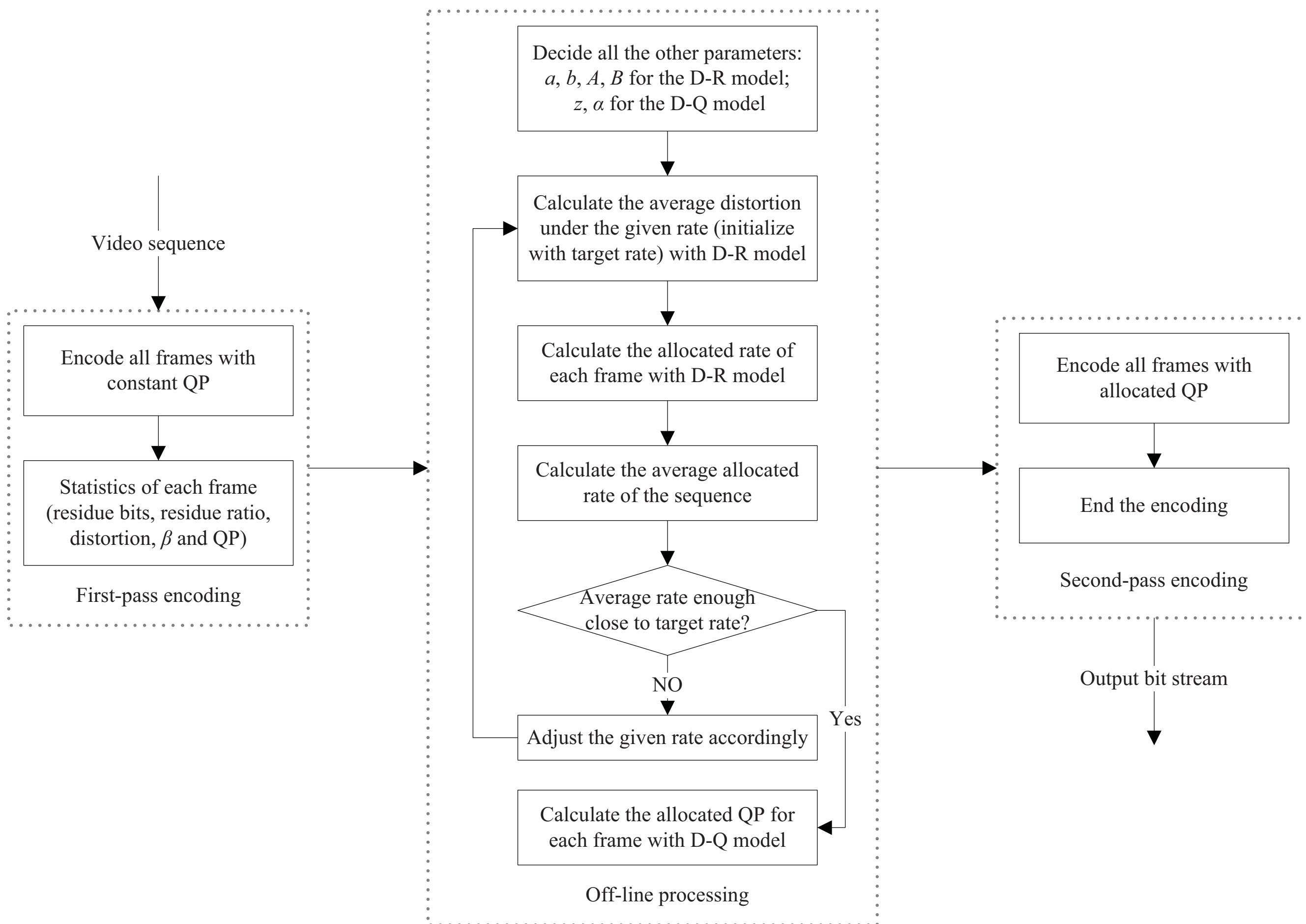


Fig. 6. Flowchart of the proposed two-pass rate control method.

the empirical values, the difference between the actual and calculated D-R data is very small. The proposed empirical values of a and b are derived base on the test sequences *Akiyo*, *Mobile*, *Silent* and *Waterfall* mentioned in Section III, specifying the selection strategy of a and b for the proposed D-R model in practical video applications.

In the two-pass rate control design for H.264 VBR coding, with the entropy coded residue bits, MSE and the parameter B computed in the first-pass encoding, the empirical values of a and b are applied to calculate the parameter A of each video frame base on the proposed D-R model. In this way, all the four parameters are accurately initialized, and in the off-line processing the proposed D-R model is utilized to predict the distortion of each frame under the target bit rate.

V. Simple Two-Pass Rate Control Method for H.264 VBR Coding

Aiming at obtaining the expected constant reconstruction quality $\overline{D}$ under the average target coding rate $\overline{R}$, the essence of the VBR rate control method is the appropriate bit allocation according to the properties of different video frames, which is represented by the calculating function $C(D)$ satisfying

$$C(\overline{D}) = \frac{1}{N}\sum_{i=0}^{N-1} R_i(\overline{D}) = \overline{R} \tag{5}$$

where $R_i\ (\overline{D})$ denotes the allocated bits of frame i. In practical implementation, the equation (5) is achieved by selecting proper QP value for each frame so that the actual average rate obtained from encoding can precisely approximate the target rate $\overline{R}$ while all the frames are of similar reconstruction distortion. Supported by the source model validation in Section III and the efficient parameter estimation strategies developed in Section IV, we propose a novel two-pass rate control method completely base on our D-R and D-Q models. In this rate control design, the D-R model is used to calculate the expected distortion $\overline{D}$ for all frames from target rate $\overline{R}$, and the D-Q model decides each frame's QP value according to $\overline{D}$. Therefore, the rate control accuracy is affected by both D-R and D-Q models while the smoothness of the reconstructed video quality is only determined by the D-Q model. The whole method is illustrated in the flowchart in Fig. 6, which mainly consists of two parts: parameter estimation and bit allocation. These two parts are performed in three stages, in the order of the first-pass encoding, the off-line processing and the second-pass encoding, which are distinguished in Fig. 6 by the dashed-line boxes. In the following subsections, the proposed rate control method is introduced from the perspective of these three stages.

A. First-Pass Encoding

In the first-pass encoding, all the frames are first encoded with a constant QP value. There is no explicit limitation for choosing the QP value. However, to better represent the practical video applications, for different formats of sequences, such as CIF (352 × 288 pixel resolution), 4CIF (704 × 576 pixel resolution), 720p (1280 × 720 pixel resolution) and 1080p (1920 × 1080 pixel resolution), we give different suggestions of the representative QP ranges. The QP is recommended to be from 36 to 24 for CIF, 4CIF and 720p sequences, and 30 to 16 for 1080p sequences, with the purpose that the average rate of the coded frames is close to the target rate $\overline{R}$.

In our two-pass rate control design, the efficient parameter estimation strategies we developed in Section IV-C and IV-D for the proposed D-Q and D-R models are performed within the first-pass encoding and off-line processing.In the first-pass encoding, the statistics of each frame i include the residue bits $R_{\text{residue\_i}}$, the total bits R_i, the distortion D_i (either under MSE or PSNR criterion), quantization parameter QP_i, the estimated standard deviation β_i of GGD and the residue transform coefficient variance B_i (equivalent to the square of β_i). Since the proposed D-R and D-Q models are both models of the residue data, for accuracy it is necessary to distinguish the entropy coded bits of quantized residue transform coefficients from the total coding bits, because the total bits involve other information like motion vectors, MB types, reference frames and so on, which should not be concerned by the D-R model. Thus we compute the ratio of the residue bits to the total bits as ratio_i for each frame i. Based on the premise that the first-pass and second-pass encodings achieve similar average coding rates, we employ ratio_i to derive the residue part of the target rate, which is the real rate utilized by the D-R model in our method.

B. Off-Line Processing

In the off-line processing, all the parameters of the proposed D-R and D-Q models that cannot be directly obtained in the first-pass encoding are finally determined. For the D-R model, the parameter a_i and b_i are selected from Table V according to the type of frame i, and then the parameter A_i is calculated with $R_{\text{residue\_i}}$, D_i, a_i, b_i and B_i using equation (2). For the D-Q model, the dead-zone ratio z is set to keep consistent with H.264/AVC standard (2/3 for I-frames and 5/6 for P- and B-frames), and the GGD shape parameter α_i is derived from equation (3) with $R_{\text{residue\_i}}$, QP_i, D_i and the estimated β_i.

As is discussed in Section IV-C, for the D-Q model, the β_i derived from the BGPE method is generally close to the optimal value obtained by data fitting. However, in frames especially where large numbers of intra MB exist, the propagation of distortion may lead to growing statistical error when the BGPE method is used to obtain β_i. In order to further guarantee the effectiveness of implementing the D-Q model, a simple scene change detection and parameter examining scheme is utilized. Since all the frames are coded by the same QP in the first-pass encoding, the frames in the same scene with the same type should have similar distortion. Under the MSE measure, a scene change frame is detected if either of the conditions below is satisfied:

$$\Delta D_i > \eta \Delta D_{\text{Avg}} \text{ and } \Delta D_i > \mu D_{Avg} \tag{6}$$

or

$$\left| D_i - D_{\text{Avg}} \right| > \kappa D_{\text{Avg}} \tag{7}$$

where ΔD_i is the MSE difference of frame i and its prior frame with the same coding type, ΔD_{Avg} is the average MSE difference of all the adjacent frames with the same coding type in the current scene, D_{Avg} is the average MSE of all the frames with the same coding type in the current scene, and η, μ, κ are all constants with empirical values of 7.0, 0.1 and 0.2, respectively. If the values of α_i and β_i are apparently abnormal, typically with the ratio greater than 2.0 or less than 0.5 to the average estimated GGD parameters of the frames with the same type in the same scene, the average α and β are used to replace α_i and β_i of frame i. The effectiveness of the model parameter estimation strategies and the parameter examining scheme proposed above are evaluated through the experimental results of the two-pass rate control method for H.264 VBR coding.

Finally, with all the model parameters successively initialized, under the PSNR distortion measurement, the target QP values for the second-pass encoding is calculated by the proposed simple bit allocation algorithm as follows.

1) Calculate the target residue bits of each frame i under the average target rate $\overline{R}$:

$$\overline{R_i} = \overline{R} \times ratio_i. \tag{8}$$

2) Calculate the average distortion $\overline{D}$ under each $\overline{R_i}$ base on the proposed D-R model

$$\overline{D} = \frac{1}{N} \sum_{i=0}^{N-1} \text{PSNR}_\text{i}(\overline{\text{R}_\text{i}}). \tag{9}$$

3) Calculate the initial allocated bit of each frame i:

$$init\_bit_i = R_i(\overline{D}) \tag{10}$$

where $R_i(\overline{D})$ is the inverse function of (3) for frame i.

4) Calculate the average allocated rate $\overline{R_{\text{alloc}}}$ under the distortion $\overline{D}$ with each $init\_\text{bit}_i$:

$$\overline{R_{\text{alloc}}} = \frac{1}{N}\sum_{i=0}^{N-1} \text{init\_bit}_\text{i}. \tag{11}$$

5) Check whether $\overline{R_{\text{alloc}}}$ is within the acceptable error range (typically 5%) of $\overline{R}$. If true, continue to the next step, otherwise update $\overline{R}$ with the following expression and repeat step 1) to 5) until the desirable precision is achieved:

$$\overline{R} = (\overline{R})^2/\overline{R_{\text{alloc}}}. \tag{12}$$

6) Calculate the $target\_Qstep_i$ and the $target\_QP_i$ of each frame i under the average distortion $\overline{D}$:

$$\text{target\_}Qstep_i = Q_i(\overline{D}) \tag{13}$$

and

$$\text{target\_}QP_i = \text{round}(6\log_2 \text{target\_}Qstep_i + 4) \tag{14}$$

where $Q_i(\overline{D})$ is the inverse function of (2) for frame i.

It should be noted that the entire bit allocation algorithm is highly stable and effective. And the iteration of the step 1) to 5) also provides a mechanism to further refine the rate control accuracy, which has the advantage of fast convergence with similar efficiency as the steepest descent method [25] for the one-dimensional optimization problem. On average, the computation complexity of the proposed bit allocation algorithm is $O(N)$, which is ignorable compared with the coding complexity.

C. Second-Pass Encoding

For the second-pass encoding, all the encoder settings remain the same as in the first-pass encoding. It should be noted that the QP_i is again utilized in the RDO process of the second-pass encoding, in order to maintain similar source statistics as in the first-pass encoding. Then the group of $target\_QP_i$ values derived in off-line processing are employed in the final coding of the residue transform coefficients. In this way, constant visual quality can be achieved under the expected target bit rate.

VI. Experimental Results

The last experiments of this paper are the implementation of the proposed two-pass rate control method over the encoder of H.264/AVC reference software JM 16.0, which are the final procedure to ultimately confirm the effectiveness of the proposed source model. For comparison, we consider other two rate-control methods: 1) the advanced rate control method of JM 16.0 and 2) Zhang *et al.*'s two-pass rate control method [4] for H.264/AVC high definition video coding.

In order to comprehensively evaluate and represent the performance of the three rate control methods, eight standard color video sequences are divided into four groups according to the resolution. For each group, two different target rates, low and high, are set to observe the adaptability of the rate control methods to different application situations.

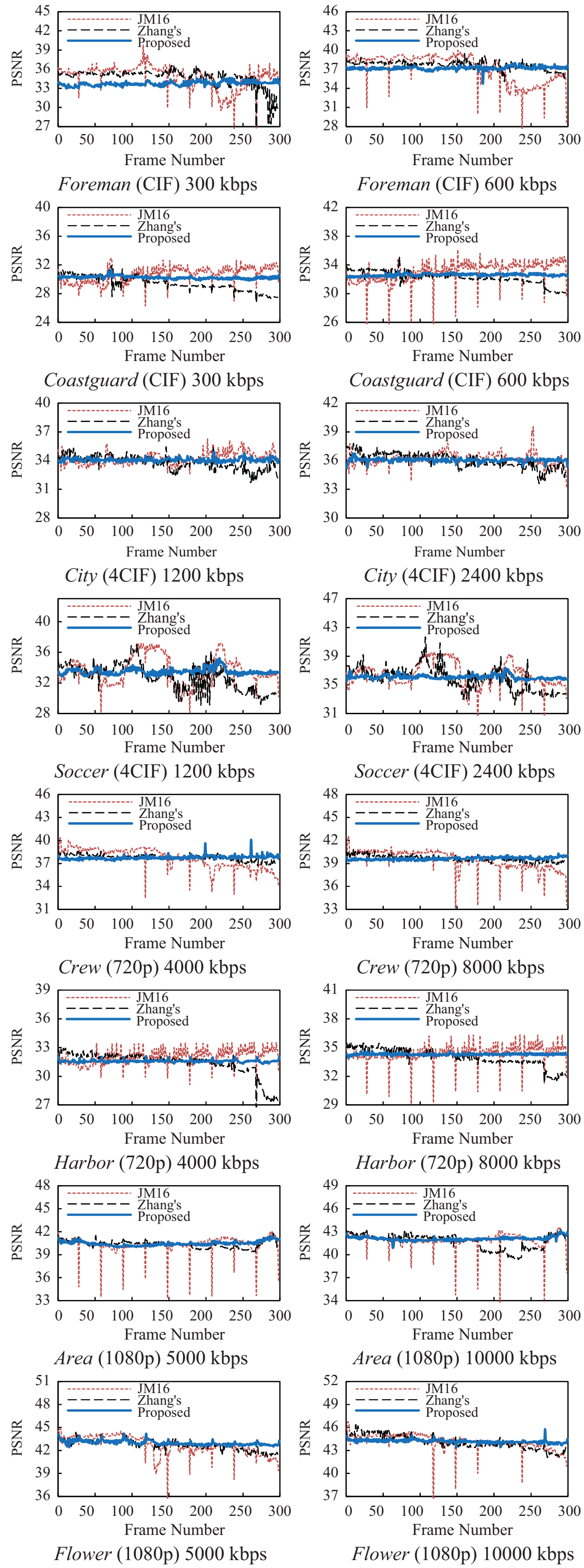


Fig. 7. Comparison of JM 16.0's advanced rate control method, Zhang *et al.* method, and the proposed method in different test sequences with various resolutions and target rates.

To be specific, the first video group consists of two CIF sequences: *Foreman* and *Coastguard*, both encoding at frame

TABLE VI
PERFORMANCE OF THREE RATE CONTROL METHODS IN TERMS OF PSNR AND OUTPUT RATE

Sequence	Target Rate (kbits/s)	Method	PSNR (dB)				Output Rate (kbits/s)
			Min	Max	Average	Variance	
Foreman (CIF)	300	JM 16.0	26.634	39.796	34.847	3.3277	308.79
		Zhang's	22.707	36.735	34.659	2.5176	309.45
		Proposed	32.995	34.639	33.741	0.1217	304.59
	600	JM 16.0	26.634	39.976	37.227	5.1768	603.59
		Zhang's	35.037	39.446	37.635	0.4071	614.86
		Proposed	34.727	37.959	37.160	0.0776	601.26
Coastguard (CIF)	300	JM 16.0	26.296	32.958	30.636	1.1969	300.50
		Zhang's	26.702	31.875	29.426	1.0756	297.91
		Proposed	29.835	31.299	30.239	0.0427	311.93
	600	JM 16.0	24.654	36.126	33.030	2.5058	598.86
		Zhang's	29.986	35.160	32.247	0.8067	598.26
		Proposed	32.201	33.167	32.602	0.0313	608.17
City (4CIF)	1200	JM 16.0	32.608	36.275	34.246	0.4125	1218.16
		Zhang's	31.723	35.444	33.802	0.5244	1204.56
		Proposed	33.621	34.904	34.068	0.0392	1205.06
	2400	JM 16.0	33.185	39.482	36.230	0.4415	2420.36
		Zhang's	33.627	37.788	36.023	0.6162	2415.66
		Proposed	35.211	36.728	36.047	0.0355	2363.91
Soccer (4CIF)	1200	JM 16.0	28.125	37.203	33.692	3.4733	1201.07
		Zhang's	29.093	37.123	32.815	2.9701	1203.70
		Proposed	32.785	35.255	33.480	0.1648	1211.61
	2400	JM 16.0	28.125	39.567	36.477	3.0639	2406.80
		Zhang's	32.211	41.628	35.883	2.7289	2426.81
		Proposed	35.401	37.390	36.088	0.0916	2346.64
Crew (720 p)	4000	JM 16.0	32.467	40.363	37.647	1.7500	4016.96
		Zhang's	36.716	38.755	37.812	0.1650	4038.76
		Proposed	37.299	40.104	37.758	0.1108	4140.81
	8000	JM 16.0	33.122	42.544	39.455	1.6410	8017.30
		Zhang's	38.655	40.831	39.635	0.1760	8072.33
		Proposed	39.199	40.287	39.644	0.0326	8239.20
Harbor (720 p)	4000	JM 16.0	29.126	33.565	32.030	0.6283	4018.85
		Zhang's	25.915	32.928	31.393	1.5629	3977.19
		Proposed	31.298	32.098	31.572	0.0147	3991.11
	8000	JM 16.0	29.101	36.333	34.385	0.9218	8002.67
		Zhang's	31.268	35.594	33.987	0.8082	8009.10
		Proposed	33.903	34.657	34.291	0.0097	8217.64
Pedestrian_area (1080 p)	5000	JM 16.0	33.281	42.110	40.287	1.4638	5011.96
		Zhang's	39.398	41.805	40.381	0.2266	5099.54
		Proposed	39.981	41.417	40.452	0.0777	4931.08
	10000	JM 16.0	33.606	43.615	41.935	1.1954	10024.19
		Zhang's	39.362	43.318	41.743	0.9557	10053.56
		Proposed	40.909	43.173	42.078	0.0650	10087.44
Sunflower (1080 p)	5000	JM 16.0	35.470	44.788	42.671	1.5655	5072.84
		Zhang's	41.076	44.379	42.792	0.4688	4986.80
		Proposed	42.427	44.185	42.985	0.0875	5083.52
	10000	JM 16.0	34.697	46.724	44.136	1.2490	10113.29
		Zhang's	42.093	46.388	44.097	0.8133	9950.24
		Proposed	43.420	45.772	44.183	0.0511	9620.54

rate 30 frames/s with target bit rate of 300 kbits/s and 600 kbits/s; the second group consists of two 4CIF sequences: *City* and *Soccer*, both encoding at frame rate 30 frames/s with target bit rate of 1200 kbits/s and 2400 kbits/s; the third group consists of two 720p sequences: *Crew* and *Harbor*, both encoding at frame rate 60 frames/s with target bit rate of 4000 kbits/s and 8000 kbits/s; the fourth group consists of two 1080p sequences: *Pedestrian_area* and *Sunflower*, both encoding at frame rate 24 frames/s with target bit rate of 5000 kbits/s and 10000 kbits/s. All of these sequences are chroma format 4:2:0 and 300 frames, covering various motion types.

In the experiments, the same encoder configuration setting as is introduced in Section III is employed in the three rate control methods. In particular, for Zhang *et al.*'s method the GOP size is set to 30. For JM 16.0, Rater Control Enable is set to 1 and RC Update Mode is set to 3 (highest level) which means the traditional method of [26] and the hybrid quadratic rate control method for I- and B- frame using bit rate statistics are both utilized, indicating a quite sophisticated scheme, while the other settings for rate control remain as default.

The PSNR comparison of the three methods is illustrated in Fig. 7 for all the test sequences, where the consistency of the reconstructed video quality is represented by the smoothness of the PSNR curve. From Fig. 7 it is explicit that in all circumstances the proposed method can almost approximate the constant visual quality while the other two methods both cause some fluctuations. In Table VI, detailed experimental results of the three methods are provided, including the statistics of PSNR and the actual output rate under each given target rate constraint.According to Table VI, the average PSNR variance of each rate control method over all the eight sequences is calculated as follows. The advanced rate control method of JM 16.0 achieves average PSNR variance of 1.8758 dB, while Zhang *et al.*'s method achieves average PSNR variance of 1.0514 dB. In contrast, the average PSNR variance of the proposed method is only 0.0658 dB, promising extremely smooth reconstructed video quality. To more comprehensively present the effectiveness of the proposed rate control method, we further compare with the Zhang *et al.*'s method in terms of PSNR variance reduction, absolute rate control error and average PSNR loss against JM 16.0's method, denoted by ΔVar, ΔRate and ΔPSNR correspondingly in Table VII, among which ΔVar is defined as

$$\Delta \text{Var} = \frac{\text{variance}_{\text{spec}} - \text{variance}_{\text{JM}}}{\text{variance}_{\text{JM}}} \tag{15}$$

where the variance$_{JM}$ and variance$_{\text{spec}}$ indicate the PSNR variance of JM 16.0's method and the specified method respectively. From Table VII, it is shown that for both of the proposed and Zhang *et al.*'s methods, the average PSNR loss compared with the JM16.0's method is ignorable, while the proposed method achieves on average about 96% reduction of JM's PSNR variance which greatly outperforms Zhang *et al.*'s average reduction of 25.5%. Meanwhile, the average rate control error of the proposed method is 1.83% compared with 0.93% of Zhang *et al.*'s method, both can well meet the demand of practical applications. The superior rate control accuracy of Zhang *et al.*'s method originates from the flexible GOP level QP adjustment mechanism [4] in the second-pass encoding. However, Zhang *et al.*'s methods till suffers from evident PSNR variance, making it less applicable in VBR coding where the smoothness of video quality is strongly required. Since Zhang *et al.*'s method implements the linear D-Q model which is only applicable in high bit rate, the bit allocation in prior GOPs often exceeds the total bits budget, resulting in drastic PSNR drop in the last GOP of the sequence, as is shown in *Foreman*, *Coastguard*, *Harbor* and *Sunflower* in Fig. 7. Moreover, based on their linear D-Q model, the

TABLE VII
RATE CONTROL PERFORMANCE COMPARISON WITH ZHANG'S METHOD IN [4]

Sequence	Target rate (kbps)	Method	ΔVar (%)	ΔRate (%)	ΔPSNR (dB)
Coastguard (CIF)	300	Zhang's	−24.34	3.15	−0.188
		Proposed	−96.34	1.53	−1.105
	600	Zhang's	−92.14	2.48	0.407
		Proposed	−98.50	0.21	−0.067
Foreman (CIF)	300	Zhang's	-10.14	0.70	-1.210
		Proposed	−96.43	3.98	−0.397
	600	Zhang's	-67.81	0.29	-0.783
		Proposed	−98.75	1.36	−0.429
City (4CIF)	1200	Zhang's	27.15	0.38	-0.443
		Proposed	−90.48	0.42	−0.178
	2400	Zhang's	39.57	0.65	−0.207
		Proposed	−91.96	1.50	−0.184
Soccer (4CIF)	1200	Zhang's	−14.49	0.31	−0.876
		Proposed	-95.25	0.97	-0.212
	2400	Zhang's	−10.93	1.12	−0.594
		Proposed	−97.01	2.22	−0.389
Crew (720 p)	4000	Zhang's	−91.46	0.97	0.165
		Proposed	−96.06	3.52	0.111
	8000	Zhang's	−89.27	0.90	0.180
		Proposed	−98.01	2.99	0.189
Harbor (720 p)	4000	Zhang's	148.76	0.57	−0.637
		Proposed	−97.66	0.22	−0.458
	8000	Zhang's	−12.33	0.11	−0.398
		Proposed	−98.94	2.72	−0.094
Area (1080 p)	5000	Zhang's	−84.52	1.99	0.095
		Proposed	−94.69	1.38	0.165
	10000	Zhang's	−20.05	0.54	−0.191
		Proposed	−94.56	0.87	0.144
Flower (1080 p)	5000	Zhang's	−70.05	0.26	0.121
		Proposed	-94.41	1.67	0.315
	10000	Zhang's	−34.89	0.50	−0.039
		Proposed	−95.91	3.79	0.047
Average		Zhang's	−25.43	0.93	−0.287
		Proposed	−95.94	1.83	−0.159

GOP level bit allocation strategy of Zhang *et al.*'s method in adjacent GOPs may cause significant video quality variation, as is shown in *Coastguard*, *City*, *Harbor* and *Pedestrian_area* in Fig. 7, where obvious fluctuation of PSNR in adjacent GOPs can be observed. As a result, the PSNR variance of Zhang *et al.*'s method is even higher than JM's results in some sequences according to Table VI.

Among all the experimental results of the proposed two-pass rate control method exhibited in Fig. 7, a few frames are still with mutations of PSNR: frame 183 of *Foreman* sequence (600 kbits/s), frame 198 and 260 of *Crew* sequence (4000 kbits/s), frame 268 of *Sunflower* sequence (10000 kbits/s). And the fluctuation of PSNR in *Soccer* sequence (1200 kbits/s) is comparatively more noticeable. These abnormal phenomena are again caused by the estimated

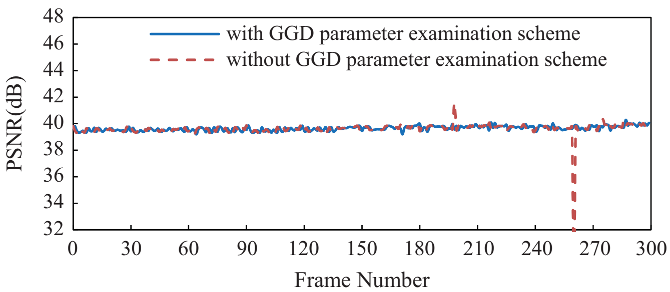


Fig. 8. Coding results of the proposed rate control method with and without the GGD parameter examination scheme for the sequence of *Crew* (720 p) under the rate constraint of 8000 kb/s.

GGD parameters which are of relatively large deviation from the optimal values. In fact, with the simple scene change detection and GGD parameter examining scheme introduced in Section V-B, the estimation error of GGD parameters has been further reduced in some extent. Fig. 8 compares the implementation of the proposed rate control method in *Crew* sequence (8000 kbits/s) with and without the parameter examining scheme. In the latter case, completely due to the deviation of the estimated β in frame 198 and 260, the obvious PSNR fluctuations in these two frames are observed, while in the former case this error is fixed by using the average values of α and β as a substitution. However, in order not to bring new error, the correcting of α and β is not applied when the scene change of a sequence is frequent, such as the *Soccer* sequence, in which case the variation of estimated GGD parameters between adjacent same type frames may also be drastic and significant. To sum up, for the problems that still exist in implementing the D-Q model, future work is required to better exploit the source properties and improve the GGD parameter estimation strategies, so that the proposed rate-control method can provide exactly constant video quality under the target rate constraint in practical video coding systems.

VII. Conclusion

In the first part of this paper, based on the comprehensive R-D analysis of DZ+UTSQ/NURQ for GGD source, we derive a source model that describes the relationship between bits, distortion and quantization step for transform coding, including the R-Q, D-Q and D-R models. The superior of the R-Q and D-Q models in predicting the rate and distortion under different QP is verified through simulation experiments.

In the second part of this paper, the effectiveness of the proposed R-Q and D-Q models is further validated base on the coding data of JM 16.0. Motivated by the observations, efficient parameter estimation strategies are developed for the implementation of the D-Q and D-R models in two-pass VBR video coding. In the end, completely based on the D-R and D-Q models, a novel two-pass rate control method is developed for H.264 VBR coding. The proposed rate control method consists of two parts: parameter estimation and bit allocation. And with the simple scene change detection and parameter examining scheme introduced, the GGD parameter estimation is further improved. Extensive experiments show that the proposed two-pass rate control method achieves average PSNR variance of 0.0658 dB compared to JM 16.0's advanced rate control method of 1.8758‘dB with ignorable coding performance degradation. Meanwhile, the average rate control error is 1.83% which is desirable for practical video applications.

Acknowledgment

The authors would like to thank Dr. D. Zhang for helping with the two-pass H.264 variable bit rate coding experiments, and all the reviewers for helping to improve the technical presentation of this paper. The authors would also like to thank www.strongene.com for their support.

References

[1] *Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification*, ITU-T Standard Rec.H.264jISO/IEC 14496-10, Mar. 2003.

[2] W.-N. Lie, C.-F. Chen, and T. C.-I. Lin, "Two-pass rate-distortion optimized rate control technique for H.264/AVC video," *Proc. SPIE, Visual Commun. Image Process.*, vol. 5960, pp. 1061–1070, Jul. 2005.

[3] D.-K. Kwon, M.-Y. Shen, and C.-C. J. Kuo, "Rate control for H.264 video with enhanced rate and distortion models," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 5, pp. 517–529, May 2007.

[4] D. Zhang, K. N. Ngan, and Z. Chen, "A two-pass rate control algorithm for H.264/AVC high definition video coding," *Signal Process., Image Commun.*, vol. 24, no. 5, pp. 357–367, Mar. 2009.

[5] D. Zhang, Z. Chen, and K. N. Ngan, "Two-pass rate control for constant quality H.264/AVC high definition video coding," in *Proc. IEEE Picture Coding Symp.*, Lisbon, Portugal, Nov. 2007, pp. 1–4.

[6] P. H. Westerink, R. Rajagopalan, and C. A. Gonzales, "Two-pass MPEG-2 variable-bit-rate encoding," *IBM J. Res. Develop.*, vol. 43, no. 4, pp. 471–488, Jul. 1999.

[7] Y. Yu, J. Zhou, Y. Wang, and C. W. Chen, "A novel two-pass VBR coding algorithm for fixed-size storage application," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 3, pp. 345–356, Mar. 2001.

[8] C. Que, G. Chen, and J. Liu, "An efficient two-pass VBR encoding algorithm for H.264," in *Proc. IEEE Int. Conf. Commun. Circuits. Syst.*, Guilin, China, Jun. 2006, pp. 118–122.

[9] *MPEG Test Model 5*, ISO/IES Standard JTC/SC29/WG11, Apr. 1993.

[10] J. Huang, J. Sun, and W. Gao, "A novel two-pass VBR coding algorithm for the H.264/AVC video coder based on a new analytical R-D model," in *Proc. IEEE Picture Coding Symp.*, Lisbon, Portugal, Nov. 2007.

[11] *Joint Video Team (JVT) Reference Software*. (2011, Jan.) [Online]. Available: http://iphome.hhi.de/suehring/tml/download/old_jm/

[12] G. J. Sullivan and S. Sun, "On dead-zone plus uniform threshold scalar quantization," *Proc. SPIE, Visual Commun. Image Process.*, vol. 5960, pp. 1041–1052, Jul. 2005.

[13] F. Müller, "Distribution shape of two-dimensional DCT coefficients of natural images," *Electron. Lett.*, vol. 29, no. 22, pp. 1935–1936, Oct. 1993.

[14] J. Sun, W. Gao, D. Zhao, and Q. Huang, "Statistical model, analysis and approximation of rate-distortion function in MPEG-4 FGS videos," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 4, pp. 535–539, Apr. 2006.

[15] W. Li, "Overview of fine granularity scalability in MPEG-4 video standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 3, pp. 301–317, Mar. 2001.

[16] N. Kamaci, Y. Altunbasak, and R. M. Mersereau, "Frame bit allocation for the H.264/AVC video coder via Cauchy-density-based rate and distortion models," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 8, pp. 994–1006, Aug. 2005.

[17] 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.

[18] I. E. G. Richardson, "The baseline profile," in *H.264 and MPEG-4 Video Compression Video Coding for Next-Generation Multimedia*. Chichester, U.K.: Wiley, 2003, ch. 6.4, pp. 187–193.

[19] G. J. Sullivan and S. Sun, “On dead-zone plus uniform threshold scalar quantization,” *Proc. SPIE, Visual Commun. Image Process.*, vol. 5960, pp. 1041–1052, Jul. 2005.

[20] J. S. Armstrong and F. Collopy, “Error measures for generalizing about forecasting methods: Empirical comparisons,” *Int. J. Forecast.*, vol. 8, no. 1, pp. 69–80, Jun. 1992.

[21] A. C. Cameron and F. A. G. Windmeijer, “An R-squared measure of goodness of fit for some common nonlinear regression models,” *J. Econ.*, vol. 77, no. 2, pp. 329–342, Apr. 1997.

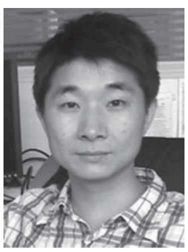
[22] K. Sharifi and A. L. Garcia, “Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, no. 1, pp. 52–56, Feb. 1995.

[23] T.-Y. Wang, H.-W. Li, Z.-M. Li, and Z.-H. Wang, “A fast parameter estimation of generalized Gaussian distribution,” in *Proc. IEEE Int. Conf. Signal Process.*, Beijing, China, Nov. 2006, pp. 101–104.

[24] X. Zhao, J. Sun, S. Ma, and W. Gao, “Novel statistical modeling, analysis and implementation of rate-distortion estimation for H.264/AVC coders,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 5, pp. 647–660, May 2010.

[25] *Steepest Descent Method*. (2012, Aug.) [Online]. Available: http://en.wikipedia.org/wiki/Steepest_descent_method

[26] Z. G. Li, F. Pan, and K. P. Lim, “Adaptive basic unit layer ratecontrol for JVT,” in *Proc. 7th Meeting JVT-G012*, Pattaya, Thailand, Mar. 2003, pp. 7–14.



Jun Sun (M’07) received the B.S. degree in computer science from University of Science and Technology Beijing, Beijing, China, in 1999, and the Ph.D. degree in computer science from Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2006.

From 2006 to 2008, he was a Post-Doctoral Researcher with the Institute of Digital Media, EECS, Peking University, Beijing. He then joined the Institute of Computer Science and Technology, Peking University, where he is currently an Associate Professor. He has published over 20 technical articles in referred journals and proceedings in the areas of video coding and analysis. His research interests include scalable video coding, rate-distortion analysis, video transcoding, video streaming, and video scheduling in wireless networks.

Yizhou Duan received the B.S. degree in computer science from Peking University, Beijing, China, in 2009, where he is currently pursuing the Ph.D. degree in computer science with the Institute of Computer Science and Technology.

His current research interests include video coding, video processing, scalable video scheduling, and rate-distortion analysis.

Jiangtao Li received the B.S. degree in electronic engineering from Sichuan University, Chengdu, China, in 2010. He is currently pursuing the Ph.D. degree in computer science with the Institute of Computer Science and Technology, Peking University, Beijing, China.

His current research interests include video coding and rate control video applications.

Jiaying Liu (S’09–M’10) received the B.E. degree from Northwestern Polytechnic University, Xi’an, China, and the Ph.D. degree (Best Graduate Hons.) Peking University, Beijing, China, in 2005 and 2010, respectively, both in computer science.

She was a Visiting Scholar with the University of Southern California, Los Angeles, from 2007 to 2008. She is currently an Associate Professor with the Institute of Computer Science and Technology, Peking University. Her current research interests include image processing, sparse signal representation, and video compression.

Zongming Guo (M’09) received the B.S. degree in mathematics and the M.S. and Ph.D. degrees in computer science from Peking University, Beijing, China, in 1987, 1990, and 1994, respectively.

He is currently a Professor with the Institute of Computer Science and Technology, Peking University. His current research interests include video coding and processing, watermarking, and communication.

Dr. Guo was a recipient of the First Prize of the State Administration of Radio Film and Television Award in 2004, the First Prize of the Ministry of Education Science and Technology Progress Award in 2006, the Second Prize of the National Science and Technology Award in 2007, the Wang Xuan News Technology Award in 2008, the Chia Tai Teaching Award in 2008, and the Government Allowance granted by the State Council in 2009. He is the Executive Member of the China-Society of Motion Picture and Television Engineers.