Delay–Power-Rate-Distortion Model for Wireless Video Communication Under Delay and Energy Constraints

Chenglin Li, *Student Member, IEEE*, Dapeng Wu, *Fellow, IEEE*, and Hongkai Xiong, *Senior Member, IEEE*

***Abstract*— Smart mobile phones are capable of performing video coding and streaming over wireless networks, but are often constrained by the end-to-end delay requirement and energy supply. To achieve optimal performance under the delay and energy constraints, in this paper we extend the traditional rate-distortion (R-D) model and the previously proposed delay R-D model to a novel delay–power-rate-distortion (d-P-R-D) model by including another two dimensions (the encoding time and encoder power consumption), which quantifies the relationship among source encoding delay, rate, distortion, and power consumption for IPPPP coding mode in H.264/AVC. We have verified the accuracy of our proposed d-P-R-D model through experiments. Based on the proposed d-P-R-D model, we develop a novel rate-control (RC) algorithm, which minimizes the encoding distortion under the constraints of rate, delay, and power. The experimental results demonstrate the superiority of the proposed RC algorithm over the existing scheme. Therefore, the d-P-R-D model and the model-based RC provide a theoretical basis and a practical guideline for the cross-layer system design and performance optimization in wireless video communication under delay and energy constraints.**

***Index Terms*— Delay–power-rate-distortion (d-P-R-D) model, H.264/AVC, rate control (RC), video coding, wireless video.**

I. Introduction

WIRELESS video communication systems, including both video encoding and streaming over wireless communication networks, have experienced extensive growth in the last decades and been used for a wide range of applications, such as video surveillance, emergency response, consumer electronics multimedia systems, and mobile video services [1]. According to the Cisco visual networking index, mobile video communication application will grow at a compound annual growth rate of 75% between 2012 and 2017, the highest growth rate of any mobile application category [2]. Such predictions lead to a natural but challenging question: how can we guarantee the quality of service (QoS) metrics, such as end-to-end distortion, and end-to-end delay, for the wireless video communication systems?

Manuscript received April 11, 2013; revised September 20, 2013; accepted December 16, 2013. Date of publication January 28, 2014; date of current version June 27, 2014. This work was supported in part by the National Science Foundation under Grants ECCS-1002214 and CNS-1116970, and in part by the National Natural Science Foundation of China under Grants U1201255, 61271218, 61228101, and 61221001. This paper was recommended by Associate Editor B. Pesquet-Popescu.

C. Li and H. Xiong are with the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: lcl1985@sjtu.edu.cn; xionghongkai@sjtu.edu.cn).

D. Wu is with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611 USA (e-mail: wu@ece.ufl.edu).

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

Digital Object Identifier 10.1109/TCSVT.2014.2302517

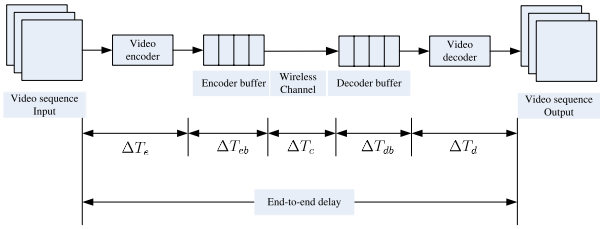


Fig. 1. End-to-end delay components of a video communication system.

From the perspective of video encoding, if the video encoder is separately investigated without consideration of its relationship with the subsequent transmission and the whole wireless video communication system, transitional rate control (RC) plays an important role that affects the overall rate-distortion (R-D) performance in the hybrid video codec design [3]. With the RD optimization (RDO), RC aims at minimizing the encoding distortion under a given constraint on the encoding rate, by appropriate selections of several coding parameters, such as quantization parameter (QP) and macroblock (MB) mode. In this case, the encoding time and power consumption are of little concern to the video encoder, since it can be assumed that there is no limit on the encoding time and power consumption.

To achieve the best end-to-end QoS performance, however, the entire cross-layer wireless video communication system is expected to appropriately assign for the video encoder both encoding time and encoding power according to the total end-to-end delay constraint and a given maximum power supply. More specifically, for a practical real-time wireless video communication system, the end-to-end delay can be broken up into several delay components that, as shown in Fig. 1, are video encoding delay ΔT_e, encoder buffer delay $\Delta T_{\rm eb}$, channel transmission delay ΔT_c, decoder buffer delay $\Delta T_{\rm db}$, and video decoding delay ΔT_d, respectively [4], [5]. In [4] and [6], the video encoding time ΔT_e and decoding time ΔT_d are both assumed to be constant. The video decoding time can be considered as a part of the video encoding time, since the encoder has to decode the video sequence as well. As illustrated in [1] and [7], however, the video encoding time (delay) is determined by the video encoding complexity, while the video encoding complexity would affect the source coding incurred distortion and bit rate. Therefore, both the distortion and bit rate of the compressed video that is transmitted over the channel are controlled by

the video encoding time. On the other hand, given an end-to-end delay constraint, if the encoding time is increased to achieve better compression performance, the allowed queuing and buffering delay at encoding/decoding buffers and channel delay for transmission will decrease accordingly, which in turn decreases the delay constrained transmission throughput and thus increases the transmission distortion of the compressed video. In general, for a given end-to-end delay constraint, the overall system performance depends on the allocation of end-to-end delay to different delay components. The change of delay assignment in one component would cause changes of the delay budget in other components, which would affect the overall system performance. Likewise, the mobile video communication system is also power limited and needs to allocate its power supply to the encoding/decoding modules and transmitting/receiving modules. For a given maximum power supply, the change of the encoding power consumption would also impact on the power budget in other modules and the overall system performance as well. Therefore, when the subsequent video transmission is considered, video encoding delay and video encoding power consumption become two new constraints which, as well, would affect the overall R-D performance of the video communication systems and need to be considered in the RC of the video encoder.

From the perspective of the cross-layer design of a wireless video communication system, the delay and power consumption constraints on the video encoder are twofolded. On the one hand, efficient video compression results in reduction of the bit rate of the video data, leading to reduction in transmission power and/or transmission delay at the physical layer, or reduction in the transmission rate and transmission error rate at data link layer. On the other hand, efficient video compression often requires high computational complexity, leading to large encoder power consumption and long encoding delay at the application layer. As implied by these two conflicting aspects, there is a tradeoff among delay d, power consumption P, rate R, and video distortion D for the design of the video encoder, which could be further applied to control the QoS performance of the entire cross-layer wireless video communication system [8]. To find the optimal tradeoff solution, we need to develop an analytic framework to model the delay–power-rate-distortion (d-P-R-D) behavior of the video encoder.

A. Related Work

Many RC schemes have been proposed in the literature to provide good video quality for the encoded video while keeping its output bit rate within the bandwidth constraint for video communication. Due to its efficiency, in the state-of-the-art video coding standard H.264/AVC, JVT-G012 [9] is adopted by simply tuning QP to meet the target bit rate. In [10], the inter-dependence between RDO and RC is further improved by QP estimation and update. Considering that the initial QP for the first I-frame would influence the RC performance, an RC scheme with adaptive QP initialization is proposed in a content-aware manner [11]. To achieve a relatively steady visual quality, [12] developed a bit allocation scheme for both I-frame and P-frame based on the frame complexity measurement and estimation model. Obviously, these schemes only focus on the R-D performance of the video encoding system, while the other two dimensions, the encoding time and power consumption, are not considered.

To formulate the RDO problem, several bit rate and quantization distortion models have been developed. Most of the existing works, e.g., [3] and [13], derive the bit rate as a function of video statistics and the quantization step size (or QP, there is a one-to-one mapping between the quantization step size Q and the QP [14], with Q increasing by 12.5% for each increment of one in QP). In addition, the quantization distortion is derived as a function of the quantization step size and video statistics for a uniform quantizer. In the R-D model of [3], both the source rate and the source distortion for a hybrid video coder with block based coding are derived as functions of the standard deviation of the transformed residuals under the assumption of Laplacian distribution. By considering the characteristics of variances of transformed residuals and compensating the mismatch between the real histogram and the assumed Laplacian distribution, [15] improved both the bit rate and distortion model where the Lagrangian-based RDO is solved by the bisection search. To achieve the optimal selection of coding parameters, [3] converted the RDO to a Lagrangian optimization problem and derived an accurate function between the single Lagrange multiplier and quantization step size. However, none of them considers the analytic model of the encoding time and power consumption, which makes the RDO not appropriate for the situation when either encoding time or encoding power is constrained. Moreover, the bisection search solver would result in a relatively high computational complexity.

Many works have been done to analyze the R-D complexity of video encoders [1], [5], [7], [16]–[18]. To derive the power-rate-distortion model for the video encoding system, [1] summarized the encoding complexity of H.263 video encoder as three modules: 1) motion estimation (ME); 2) precoding (transform, quantization, inverse quantization, and inverse transform); and 3) entropy coding. The relationship among the encoding complexity, rate, and distortion was analyzed, and the power consumption level was adopted to represent the encoding complexity. Unfortunately, this P-R-D model is dedicated to only H.263 video encoder. The model should be evolved since H.264/AVC uses the tree-structured motion compensation with seven inter-modes, which causes the ME consumes much more encoding complexity than the other two modules. As a matter of fact, [1] also fails to consider the dimension of the encoding time, which is relevant to the encoding complexity. To tackle these issues, the delay-rate-distortion (d-R-D) model of H.264/AVC video encoders was proposed and analyzed in [5] and [7] for both IPPPP and IBPBP coding modes. This d-R-D model depends on the quantization step size and the standard deviation of transformed residuals in ME, which was further fitted as functions of coding parameters in ME. However, this model did not consider and analyze the encoding power consumption that is also closely related to the encoding complexity. In addition, it neglects the critical impact of the quantization step size on both the standard deviation of transformed residuals and the

encoding delay component, which will be demonstrated based on extensive experiments in Section II.

B. Proposed Research

To the best of our knowledge, there has been no analytical framework for the d-P-R-D modeling of the video encoding system, which is of great importance to analyze the effect of the video encoding time and power consumption on the R-D performance of the video encoder. In this paper, we extend from the traditional R-D model [3], [15] and the d-R-D model previously proposed in [5] and [7], and accordingly develop an analytic framework to model, control, and optimize the d-P-R-D behavior of the H.264/AVC video encoding system. More specifically, our contributions in this paper are twofold. First, four dimensions (rate, distortion, delay, and power) that jointly determine the performance of the H.264/AVC video encoder are derived as functions of coding parameters (search range and number of reference frames in ME and quantization step size), respectively. Here, without loss of generality, the coding structure of the H.264/AVC encoder is chosen to be IPPPP coding mode, which is also reasonable since as will be introduced, the ME module for inter-coded P-frames takes the major part of the entire encoding complexity. The model accuracy has also been validated and compensated by considering the statistics of both the current frame and the previous frame. Second, the proposed d-P-R-D model is applied to formulate the source RC problem as a d-P-R-D optimization problem with respect to the search range and quantization step size. Compared with the existing work on source RC aiming at minimizing the video encoding distortion, we have introduced two more constraints corresponding to the encoding time and the encoding power, in addition to the traditional rate constraint. Furthermore, a practical algorithm based on both Karush–Kuhn–Tucker (KKT) conditions and sequential quadratic programming (SQP) methods for the d-P-R-D optimization-based RC problem is developed, which can produce both primal (search range and quantization step size) and dual (Lagrange multipliers) solutions simultaneously and efficiently in an iterative way. The proposed d-P-R-D model and model-based RC algorithm provide a theoretical basis, as well as a practical guideline, for the cross-layer system design and performance optimization in wireless video communication under delay and energy constraints. Using the proposed d-P-R-D model, we can optimize the cross-layer performance (e.g., end-to-end distortion) by appropriately allocating the delay and power budget to components within the wireless video communication system.

C. Paper Organization

The rest of this paper is organized as follows. In Section II, we derive the d-P-R-D source coding model for H.264/AVC and verify the model accuracy based on experiments. In Section III, we formulate a d-P-R-D optimization-based source RC problem, and accordingly develop a practical algorithm based on KKT conditions and SQP methods to determine the optimal selection of coding parameters. Section IV presents the experimental results, which demonstrate the accuracy of the d-P-R-D model, the convergence behavior, and performance of the proposed algorithm. The concluding remarks and the future work are given in Section V.

II. D-P-R-D Source Coding Model

According to the R-D model proposed [3], [15], both source rate and source distortion for a hybrid video coder with block-based coding, e.g., H.264/AVC encoder, are based on the distribution of transformed residuals, which is mainly determined by the ME accuracy and quantization distortion. More specifically, under the assumption that the transformed residuals in ME follow Laplacian distribution [3], [19], the source rate and distortion of an inter-coded P-frame in IPPPP coding mode can be derived as functions of the standard deviation σ of the transformed residuals and the quantization step size Q (or QP). The extension to other distributions (e.g., generalized Gaussian distribution and Cauchy distribution) is also straightforward [15], since the transform coefficients are supposed to be independent and identically distributed (i.i.d.).

To further analyze ME accuracy in H.264/AVC [20], the standard deviation σ of transformed residuals depends on the following four coding parameters: 1) MB coding mode; 2) ME search range λ; 3) the number of reference frames θ [5], [7]; and 4) quantization step size Q. If the function relationship of $\sigma(\lambda, \theta, Q)$ can be established for H.264/AVC encoder, the source rate and distortion will become functions of ME parameters λ, θ, and quantization step size Q. On the other hand, both encoding time and encoding power are monotonously increasing with encoding complexity. Since ME module is the most complexity exhausting part within the entire encoding process, it is reasonable to approximate the entire encoding complexity by ME complexity, which is determined by the quantization step size as well as the number of sum of absolute difference (SAD) operations for each MB partition (or subpartition): $\#\text{SAD} = (2\lambda + 1)^2 \times \theta$ [5], [7]. Therefore, by translating the specific coding behavior into encoding complexity, both encoding time and power can be expressed as functions of the quantization step size Q and ME parameters λ and θ. If the encoding time is equivalently considered as video encoding delay, then the entire d-P-R-D source coding model can be formulated.

A. Source Rate and Distortion Model

To develop the source rate and distortion model, the closed-form function of $\sigma(\lambda, \theta, Q)$ would be generated first. Due to the lack of any prior knowledge of the exact function form, a basic means is to draw the relationship of σ versus λ, θ, and Q, which can be fitted by a known function form [5], [7]. To achieve it, the JM18.2 [21] coding engine is tested with the IPPPP coding mode, where the *Bus* (QCIF, 176 × 144) and *Foreman* (CIF, 352 × 288) video sequences are used to collect the statistics with a wide range of scene activity pattern, including camera movement and large object motion (*Bus*), medium but complex motion (*Foreman*).[1] For a fair

[1]Due to page limitation, please refer to [22] for more modeling and experimental results on other test sequences, such as the *Mobile* sequence that contains motion with zooming effects.

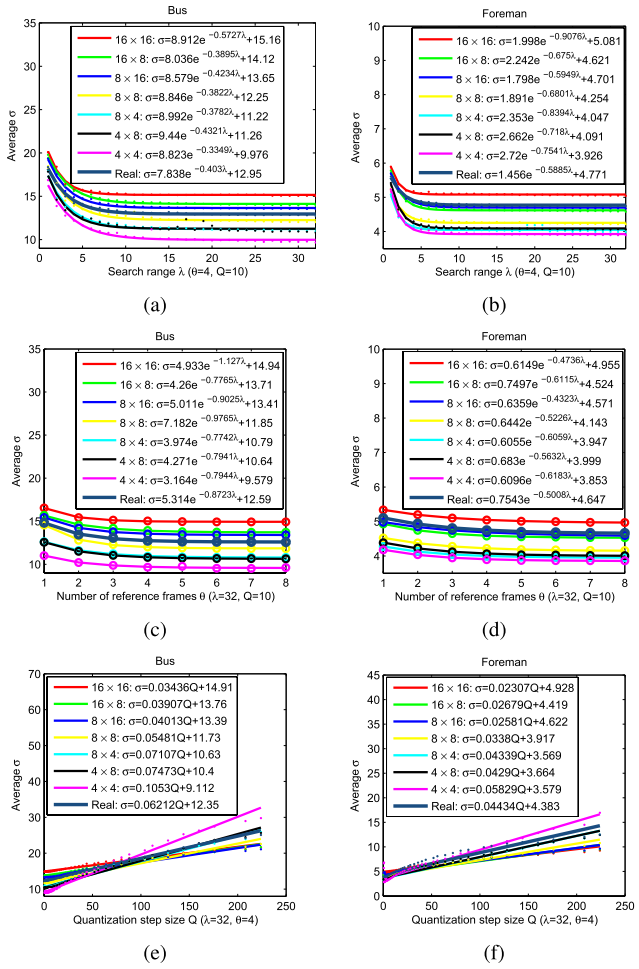


Fig. 2. Relationship and fitting results of average σ versus λ, θ, and Q. (a), (c), and (e) *Bus* sequence. (b), (d), and (f) *Foreman* sequence.

evaluation, all MBs in the experiments would select the same coding mode from the eight inter-modes except skip mode to exclude the potential influence of MB coding mode in ME. Accordingly, these inter-modes are shown by index 1–7 as in JM configuration, (i.e., assigning index 1 to 16 $\times$ 16 inter-mode, index 2 to 16 $\times$ 8 inter-mode, etc.).

Since λ, θ, and Q are independently tuned parameters in JM 18.2 configurations, we separately evaluate their impacts on the average standard deviation σ of transformed residuals. For all the seven inter-modes and the real mode selection where each MB chooses the best inter-mode based on RDO, respectively, Fig. 2(a) and (b) shows the relationship between average standard deviation σ of transformed residuals and search range λ, with fixed θ and Q. Likewise, the impact of number of reference frames θ on the average standard deviation σ of transformed residuals is shown in Fig. 2(c) and (d), at fixed λ and Q. Similarly, as in [5] and [7], it is verified that an exponential function with a constant vertical translation can be used to fit the curves in Fig. 2(a)–(d) as

$$\sigma(\lambda) = a_1 e^{-b_1 \lambda} + c_1 \tag{1}$$

$$\sigma(\theta) = a_2 e^{-b_2 \theta} + c_2 \tag{2}$$

where a_1, b_1, c_1, a_2, b_2, and c_2 are fitting parameters. When λ and θ are fixed, the relationship between average standard deviation σ of transformed residuals and quantization step size Q is shown in Fig. 2(e) and (f), which show that the curves can be simply fitted by a linear function

$$\sigma(Q) = a_3 Q + b_3 \tag{3}$$

where a_3 and b_3 are fitting parameters.

To have a better understanding of (1)–(3), we will discuss the impact of the aforementioned four different factors on σ. In general, an inter-mode with higher mode index (i.e., with smaller size of MB partitions) will lead to better prediction, and thus the standard deviation σ of transformed residuals tends to be smaller. For the real mode selection, since MB can choose a best mode out of all the seven inter-modes, the value of σ is bounded within modes 1 and 7. On the other hand, either a larger search range λ or a larger number of reference frames θ will result in a bigger 3-D search cube in ME and thus a better prediction, which would also lead to a smaller σ. The last factor, quantization step size Q, would affect the distortion of the reference frames. In general, the distortion of the reference frames will be increased by the selection of larger Q, which tends to result in larger σ.

From Fig. 2, it can also be observed that θ has a little contribution to the change of σ compared with the other two parameters. For example, for the same mode, the changing rate of σ versus θ is much smaller than that of σ versus λ and σ versus Q. In the real mode selection, we could investigate the function $\sigma(\lambda, Q)$ for given $\theta = \theta_0$, which approximates the function $\sigma(\lambda, \theta, Q)$, for the sake of simplicity. Another motivation is that by fixing the value of θ, the computational complexity for fitting the standard deviation function $\sigma(\lambda, \theta, Q)$ would be decreased. Fig. 3 shows the 2-D fitting results of function $\sigma(\lambda, Q)$ with θ fixed at one. Considering both the exponential relationship with λ and linear relationship with Q, the fitted 2-D function of σ can be represented in the form of

$$\sigma(\lambda, Q) = a e^{-b\lambda} + c + dQ \approx \sigma(\lambda, \theta, Q) \tag{4}$$

where a, b, c, and d are fitting parameters. With the fitting results, we have obtained the closed-form function of $\sigma(\lambda, Q)$, which is an approximation of $\sigma(\lambda, \theta, Q)$. To better assess the fitting performance, both R-square and root-mean-square error (RMSE) are used as metrics to measure the degree of data variation from the proposed model (4), as shown in Fig. 3.

For i.i.d. zero-mean Laplacian source under the uniform quantizer, the closed-form functions of entropy of quantized transform coefficients and quantization distortion have been

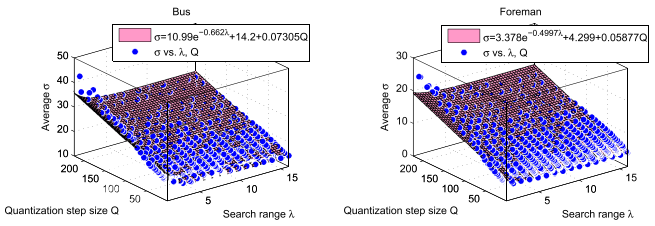


Fig. 3. 2-D fitting of average σ versus λ and Q when $\theta = 1$. (a) *Bus* sequence with R-square = 0.965 and RMSE = 0.845. (b) *Foreman* sequence with R-square = 0.936 and RMSE = 0.884.

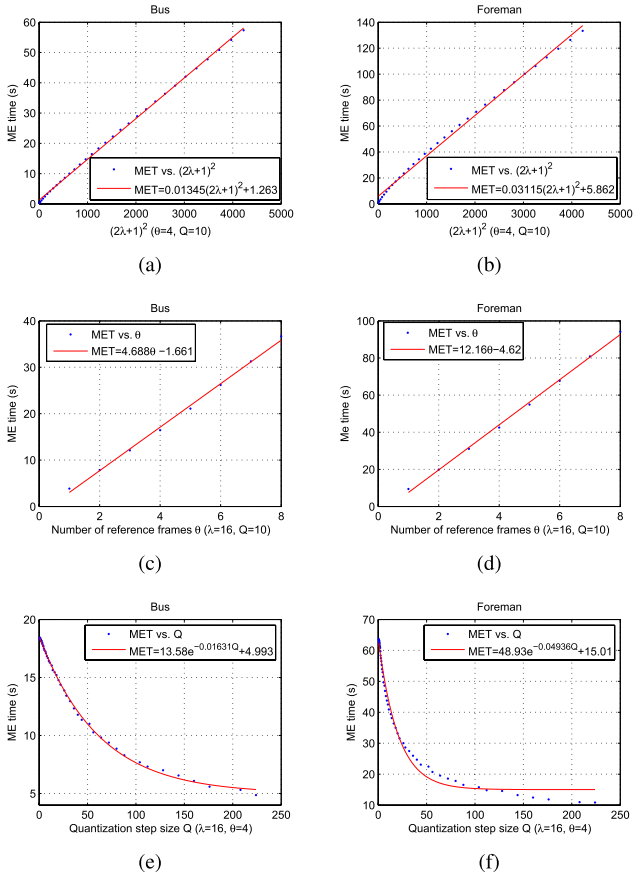


Fig. 4. Relationship and fitting results of MET versus $(2\lambda+1)^2$, θ, and Q. (a), (c), and (e) *Bus* sequence. (b), (d), and (f) *Foreman* sequence.

provided in [15]. Furthermore, the entropy of quantized transformed residual can approximate the source rate model when no effect by side information, such as MB type and motion information [3]. Hence, the closed form of source rate model is given by

$$\begin{aligned} R(\Lambda, Q) &= H(\Lambda, Q) \\ &= -P_0 \log_2 P_0 + (1-P_0)\left[\frac{\Lambda Q \log_2 e}{1-e^{-\Lambda Q}} - \log_2(1-e^{-\Lambda Q}) - \Lambda Q \gamma \log_2 e + 1\right] \end{aligned} \tag{5}$$

where $\Lambda = \sqrt{2}/\sigma$ is the Laplace parameter that is one-to-one mapping of σ, γ Q represents the rounding offset when γ is a parameter between (0, 1), such as 1/6 for H.264/AVC inter-frame coding [3], and $P_0 = 1 - e^{-\Lambda Q(1-\gamma)}$ is the probability of quantized transform coefficient being zero. Since the source distortion is only caused by quantization error, the corresponding source distortion model is expressed as

$$D(\Lambda, Q) = \frac{\Lambda Q e^{\gamma \Lambda Q}(2 + \Lambda Q - 2\gamma \Lambda Q) + 2 - 2e^{\Lambda Q}}{\Lambda^2 (1 - e^{\Lambda Q})}. \tag{6}$$

Due to page limitation, the derivation process and proof of (5) and (6) are given in [22]. Integrating (4) and $\Lambda = \sqrt{2}/\sigma$ into (5) and (6), both source rate and distortion can be further expressed as functions of λ and Q, i.e., $R(\lambda, Q)$ and $D(\lambda, Q)$.

B. Encoding Time and Power Model

As introduced in [1], the encoding complexity comprises of three segments: 1) ME; 2) precoding (transform, quantization, inverse quantization, and inverse transform); and 3) entropy coding. Theoretically, the entire encoding time is the sum of individual duration of each of the three segments. To achieve higher compression efficiency, H.264/AVC uses tree structured motion compensation with seven inter-modes, which causes ME as the most time consuming part within all the three segments of the encoder. As demonstrated in [5] and [7], it is reasonable to approximate the entire encoding time by the ME time (MET) for IPPPP coding mode. It is worth mentioning that the MET ratio throughout this paper is attained by exhaustive full search, which can guarantee achieving the optimal motion vector and minimum predictive residuals.

Technically, the MET of an inter-coded P frame can be derived as the total number of CPU clock cycles consumed by its SAD operation divided by the number of clock (Hz) [5], [7]. Namely, the encoding time for an inter-coded P frame is approximated by the MET as

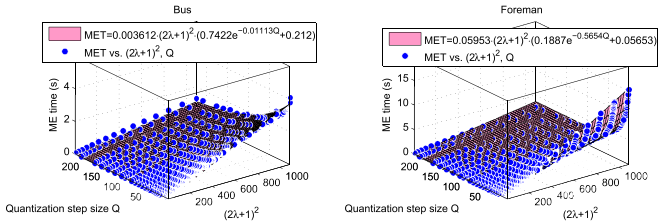
$$d(\lambda, \theta) \approx \mathrm{MET}(\lambda, \theta) = \frac{N(2\lambda+1)^2 \theta \cdot c_0}{f_{\mathrm{CLK}}} \tag{7}$$

where N is the number of MBs in a frame, $(2\lambda+1)^2\theta$ is the total number of SAD operations in a 3-D search cube for each MB, c_0 is the number of clock cycles of one SAD operation over a given CPU, and f_{CLK} is the clock frequency of the CPU. Through the dynamic voltage scaling model to control power consumption of a microprocessor [23], [24], f_{CLK} can be further related to the CPU power consumption

$$P = k \cdot f_{\mathrm{CLK}}^3 \tag{8}$$

where k is a constant in the dynamic voltage scaling model and determined by both the supply voltage and the effective switched capacitance of the circuits [25].

To justify the theoretical encoding time model in (7), Fig. 4 is provided to further investigate the relationship between the MET and ME parameters λ, θ, and quantization step size Q. It can be observed from Fig. 4(a) and (b) that MET can be fitted by a linear function of either search area $(2\lambda+1)^2$ or the number of pixels ever searched in a reference frame. Similarly, Fig. 4(c) and (d) shows that MET can also be fitted by a linear function of number of reference frames θ. In accordance with (7), the linear relationship between MET and $(2\lambda+1)^2$ or θ is obvious, since $(2\lambda+1)^2 \cdot \theta$ would form a 3-D search cube for an MB in ME and represent the total number of SAD operations per MB. In Fig. 4(e) and (f), MET is illustrated to be dependent on quantization step size Q too. The relationship can be well fitted by an exponential function with a constant translation along the MET axis. Specifically, the higher the quantization step size is, the more likely an MB would satisfy the skip mode condition, and thus the more MBs would end up with skip mode as the real coding mode, which can lower the encoding complexity. The higher the quantization step size is, the fewer number of SAD operations for each MB is involved, and the lower complexity to encode the inter-frame. Hence, (7) is modified to reflect such dependency between the MET and



(a) (b)

Fig. 5. 2-D fitting of MET versus $(2\lambda+1)^2$ and Q with $\theta = 1$ for the (a) *Bus* sequence, and (b) *Foreman* sequence.

the quantization step size

$$d(\lambda, \theta, Q) \approx \mathrm{MET}(\lambda, \theta, Q) = \frac{N(2\lambda+1)^2\theta \cdot \alpha(Q) \cdot c_0}{f_{\mathrm{CLK}}} \quad (9)$$

where $\alpha(Q)$ denotes the ratio of the actual number of SAD operations in the JM codec to the theoretical total number of SAD operations, and thus $N(2\lambda+1)^2\theta \cdot \alpha(Q)$ represents the actual number of SAD operations of a frame.

Fig. 5 shows the relationship of MET versus search area and quantization step size, with number of reference frames fixed at one, namely, the functional form of $d(\lambda, \theta, Q|\theta = 1)$. Comparing the 2-D fitting results with (9), it can be observed that $N \cdot \alpha(Q) \cdot c_0 / f_{\mathrm{CLK}} = 0.003612 \cdot (0.7422e^{-0.01113Q} + 0.212)$, where $N = 99$ for the QCIF video sequence. The correctness of (9) can also be validated by the results in Fig. 4.

C. Model Accuracy Verification

1) Source Rate and Distortion: According to [26], the transform coefficients in a video encoder are not i.i.d. As described in [15], the 16 coefficients in a 4×4 integer transform show a decreasing variance in the zigzag scan order. To improve the model accuracy of source rate and distortion, the coefficients should be modeled by random variables with different variances. The joint entropy and overall quantization distortion for the 16 coefficients can be applied to the source rate and distortion models. Specifically, suppose (x, y), $x, y \in \{0, 1, 2, 3\}$ is the position of a specific coefficient in the 2-D transform domain of the 4×4 integer transform, the variance $\sigma^2_{(x,y)}$ is derived by the average variance σ^2 of all positions as

$$\sigma^2_{(x,y)} = 2^{-(x+y)} \cdot \sigma^2_{(0,0)} = 2^{-(x+y)} \cdot \frac{1024}{225}\sigma^2. \quad (10)$$

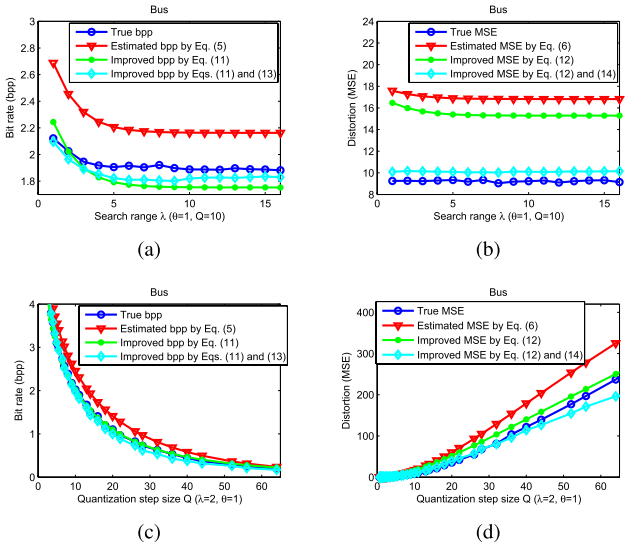
Therefore, the source rate and distortion model can be improved by

$$R(\Lambda, Q) = H(\Lambda, Q) = \frac{1}{16}\sum_{x=0}^{3}\sum_{y=0}^{3} H(\Lambda, Q)_{(x,y)} \quad (11)$$

$$D(\Lambda, Q) = \frac{1}{16}\sum_{x=0}^{3}\sum_{y=0}^{3} D(\Lambda, Q)_{(x,y)} \quad (12)$$

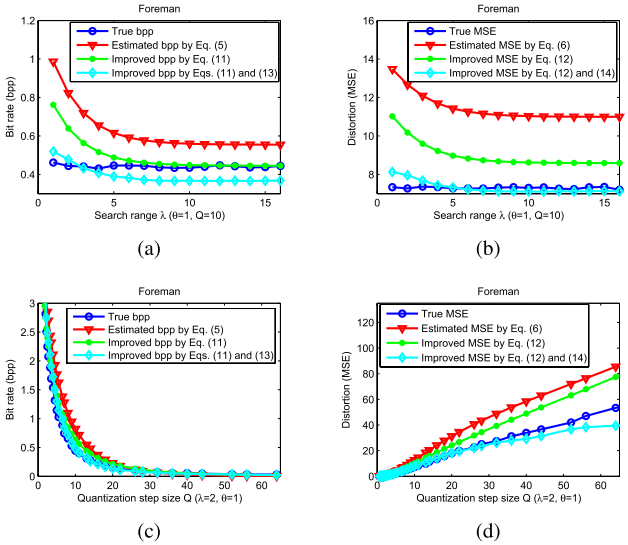
where $H(\Lambda, Q)_{(x,y)}$ and $D(\Lambda, Q)_{(x,y)}$ are the entropy and distortion associated with coefficient in (x, y), respectively.

Considering the Laplacian distribution representing the residual probability distribution may deviate significantly from



(a) (b) (c) (d)

Fig. 6. Compensation results of (a) and (c) source rate model, and (b) and (d) distortion model for the *Bus* sequence.



(a) (b) (c) (d)

Fig. 7. Compensation results of (a) and (c) source rate model, and (b) and (d) distortion model for the *Foreman* sequence.

the residual histogram, in addition, the mismatch would be compensated by the statistics from the previous frame [5], [15]

$$R_t^k = \frac{R_t^{k-1} R_l^k}{R_l^{k-1}} \quad (13)$$

$$D_t^k = \frac{D_t^{k-1} D_l^k}{D_l^{k-1}} \quad (14)$$

where superscripts $k-1$ and k denote the frame index of two consecutive frames and subscripts l and t denote the estimated value under Laplacian distribution assumption and the true value, respectively.

In Figs. 6 and 7, the accuracy of the proposed source rate model (5) and distortion model (6) is evaluated in comparison with the actual bits per pixel and MSE measures. In addition, the compensated model estimate of source rate and distortion

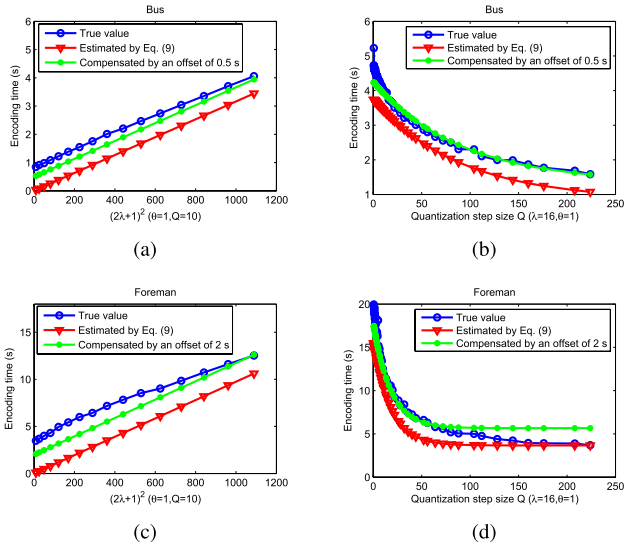


Fig. 8. Compensation results of encoding time model. (a) and (b) *Bus* sequence. (c) and (d) *Foreman* sequence.

on (11) and (12), as well as the improvement on (13) and (14), are illustrated.

2) Encoding Time: Although the entire encoding time can be approximated by the ME segment, we would modify encoding time model slightly by compensating an offset to the precoding and entropy coding segments, which is much smaller compared with MET. In experiments, the off-set time of the current frame is set as the average difference time between actual encoding time and model-based encoding time estimation of previous frames within a sliding window. Fig. 8 shows the true encoding time and the estimated encoding time, which can be observed that the difference between the encoding time model (9) and its true value is relatively small. As the encoding time increases, the difference would become less significant.

III. D-P-R-D Optimization-Based Source RC and Algorithm Design

In this section, we apply the proposed models in Section II to formulate the source RC problem as a d-P-R-D optimization with respect to search range λ and quantization step size Q, and accordingly design a practical RC algorithm.

A. d-P-R-D Optimization-Based Source RC

In the previous section, we have derived the analytical models of rate, distortion, delay, and power, as functions of three parameters, search range λ, number of reference frames θ, and quantization step size Q, respectively. As discussed in Section II-A, however, θ is a less effective parameter on the change of σ than the other two parameters. To this end, we keep the value of θ fixed and choose λ and Q as the tuning parameters, and thus the d-P-R-D optimization-based source RC problem can be formulated as

$$\begin{aligned} \min \quad & D(\lambda, Q) && (15a)\\ \text{s.t} \quad & R(\lambda, Q) \le R_{\max} && (15b)\\ & d(\lambda, Q) \le d_{\max} && (15c)\\ & P \le P_{\max}. && (15d) \end{aligned}$$

Compared with the existing works on source RC problems, the optimization problem (15) is constrained by two more conditions of the encoding time and encoding power, in addition to the rate. Ideally, an efficient video encoder is expected to encode a raw video sequence into a bit stream with minimum distortion, rate, encoding delay, and encoding power. From the analysis of the proposed d-P-R-D models in Section II, it can be observed that a larger search range λ as well as a smaller quantization step size Q is required to achieve the objective of minimizing the distortion $D(\lambda, Q)$. On the other hand, however, the decrement of Q will cause the rate $R(\lambda, Q)$ to increase, and the encoding delay $d(\lambda, Q)$ will also become greater with either λ increasing or Q decreasing. Furthermore, for a coding parameter pair (λ, Q), the encoding delay $d(\lambda, Q)$ can be further reduced by increasing the encoding power P, while the distortion and rate are still kept at the same level. Therefore, it is infeasible for a video encoder to simultaneously achieve the goals of minimizing distortion, rate, encoding delay, and encoding power. Accordingly, the d-P-R-D optimization-based source RC problem (15) is to find the Pareto optimal tradeoff among the four optimization criteria. As a matter of fact, the target of such optimization is to minimize the distortion $D(\lambda, Q)$ for given rate $R_{\max}$, encoding delay $d_{\max}$, and encoding power $P_{\max}$, by appropriate selections of coding parameter pair (λ, Q).

Depending on the estimation accuracy of the d-P-R-D models, the source RC problem (15) can be applied to a desired coding unit, e.g., a sequence, a group of pictures (GOP), a frame, or an MB. For example, if the d-P-R-D model is applied to a stream, (15) can be regarded to solve the sequence-level RC problem. If it is applied to a frame, (15) can behave as a frame level RC. Without loss of generality, a sequence-level RC problem will be imposed on (15) with a practical solution.

B. Algorithm Design

Considering (8) and (9), for the coding parameter pair (λ, Q), the minimum encoding delay would be achieved with the maximal power $P_{\max}$. In other words, the feasible set of coding parameters (λ, Q) constrained by a given maximum encoding delay would become the largest if and only if the encoding power reaches the maximum. According to Proposition 1, the power constraint (15d) in (15) is, therefore, an active constraint at the optimal coding parameter pair (λ^*, Q^*), and the source RC problem (15) can be transformed to an equivalent problem (16).

Proposition 1: Problem (15) is equivalent to the following optimization problem:

$$\begin{aligned} \min \quad & D(\lambda, Q) && (16a)\\ \text{s.t.} \quad & R(\lambda, Q) \le R_{\max} && (16b)\\ & d(\lambda, Q) \le d_{\max} && (16c)\\ & P = P_{\max}. && (16d) \end{aligned}$$

Proof: Please refer to [27]. ■

Either the Lagrange multiplier method [28]– [30] or the dynamic programming approach [31] can solve (16). The former is preferred throughout this paper since it can

be implemented independently in each coding unit. In comparison, the dynamic programming approach requires a tree representing all possible solutions to grow over multiple coding units. The computational complexity would grow exponentially with the number of coding units, which is not affordable for practical applications. With the Lagrange multiplier method, (16) can be converted to an unconstrained problem

$$\min \quad L(\lambda, Q, \mu, \eta) = D(\lambda, Q) + \mu \cdot [R(\lambda, Q) - R_{\max}] + \eta \cdot [d(\lambda, Q) - d_{\max}] \tag{17}$$

where $\mu \geq 0$ and $\eta \geq 0$ are Lagrange multipliers associated with two inequality constraints, and the equality constraint $P = P_{\max}$ can be integrated to $d(\lambda, Q)$ as

$$d(\lambda, Q) = \frac{N(2\lambda + 1)^2 \theta \cdot \alpha(Q) \cdot c_0}{\sqrt[3]{k^{-1} P_{\max}}}. \tag{18}$$

Based on the theorem in [28], we have the following theorem that relates the solution to the unconstrained problem (17) to the solution to the constrained problem (16).

Theorem 1: For any $\mu \geq 0$, $\eta \geq 0$, the solution (λ^*, Q^*) to the unconstrained problem (17) is also the solution to the constrained problem (16) with the constraints $R(\lambda, Q) \leq R(\lambda^*, Q^*)$, and $d(\lambda, Q) \leq d(\lambda^*, Q^*)$.

Proof: Please refer to [30]. ■

It should be noted that although Theorem 1 does not guarantee any solution to the constrained problem (16), it shows that for any nonnegative μ and η, there is a corresponding constrained problem whose solution is identical to that of the unconstrained problem. Therefore, as μ and η are swept over the range $[0, +\infty]$, if there exists a specific pair of μ^* and η^*, which makes $R(\lambda^*, Q^*)$ and $d(\lambda^*, Q^*)$ happen to be equal to $R_{\max}$ and $d_{\max}$, then (λ^*, Q^*) is the desired solution to the constrained problem (16).

To estimate the corresponding μ^* and η^* in practice, the bisection search method [15], [30] is commonly used for iterations to acquire the best Lagrange multiplier. However, two disadvantages have prevented such method from being suitably applied to the unconstrained problem (17). First, the bisection search method would perform worse or even fail to get the solution when extended to 2-D search scenario. For example, if we simultaneously bisect the intervals for μ and η and then select a subinterval for each of these two Lagrange multipliers based on their own criteria, respectively, the best solution μ^* and η^* might be excluded by such independent bisections. Therefore, to get the best solution, in many cases, we have to implement an exhausting search over two dimensions, which is very time consuming. Second, even if the bisection search can suitably work for searching two Lagrange multipliers simultaneously, we still need to update the Lagrange multipliers in each iteration, and then solve the corresponding unconstrained problem (17) to get the solution. It means that the update of primal and dual variables is not synchronous and may cause high computational complexity. Another analytical way to determine the best Lagrange multiplier is the model-based method [3], [29], which focuses on RDO and accordingly derives an accurate function between the single Lagrange multiplier and Q. Without consideration of the encoding time and power, however, the derived function is no longer accurate and thus cannot be directly applied to the d-P-R-D optimization.

In the following, we propose a practical algorithm for the d-P-R-D optimization-based source RC problem (16) on the basis of KKT conditions and SQP methods, which can produce both primal (λ^* and Q^*) and dual (Lagrange multipliers μ^* and η^*) solutions simultaneously in an iterative way. To solve the first-order necessary conditions of optimality for problem (16), the SQP methods [32], [33] can be used to construct a quadratic programming (QP) subproblem at a given approximate solution, and then to employ the solution to this subproblem to construct a better approximation. This process is iterated to create a sequence of approximations that is expected to converge to the optimal solution $(\lambda^*, Q^*, \mu^*, \eta^*)$. Specifically, given an iterate $(\lambda^k, Q^k, \mu^k, \eta^k)$, a new iterate $(\lambda^{k+1}, Q^{k+1}, \mu^{k+1}, \eta^{k+1})$ can be obtained by solving a QP minimization subproblem given by

$$\min_{\delta^k} \quad \frac{\partial D(\lambda^k, Q^k)}{\partial \lambda} \cdot \delta_\lambda^k + \frac{\partial D(\lambda^k, Q^k)}{\partial Q} \cdot \delta_Q^k + \frac{1}{2} \delta^{k\,T} \cdot \nabla^2 L(\lambda^k, Q^k, \mu^k, \eta^k) \cdot \delta^k \tag{19a}$$

s.t

$$\frac{\partial R(\lambda^k, Q^k)}{\partial \lambda} \cdot \delta_\lambda^k + \frac{\partial R(\lambda^k, Q^k)}{\partial Q} \cdot \delta_Q^k + R(\lambda^k, Q^k) = R_{\max} \tag{19b}$$

$$\frac{\partial d(\lambda^k, Q^k)}{\partial \lambda} \cdot \delta_\lambda^k + \frac{\partial d(\lambda^k, Q^k)}{\partial Q} \cdot \delta_Q^k + d(\lambda^k, Q^k) = d_{\max} \tag{19c}$$

where the derivative operators ∇^2 is used to refer to the second-order Hessian matrix with respect to primal variables λ and Q, and $\delta^k = (\delta_\lambda^k, \delta_Q^k)^T = (\lambda^{k+1} - \lambda^k, Q^{k+1} - Q^k)^T$ is the vector representing the update directions of primal variables. For the derivation of (19), please refer to [22].

The aforementioned SQP algorithm, though can be used to appropriately solve (16), suffers from two deficiencies similar to the Newton's method. First, it requires at each iteration the calculation of second-order Hessian matrix $\nabla^2 L(\lambda^k, Q^k, \mu^k, \eta^k)$, which could be a costly computational burden and in addition might not be positive definite. To address this issue, we can use the quasi-Newton method instead to construct an approximate Hessian matrix B^k by which $\nabla^2 L(\lambda^k, Q^k, \mu^k, \eta^k)$ is replaced in (19). In practice, such an approximation B^k can be obtained by the Broyden–Fletcher–Goldfarb–Shanno method [34]

$$B^{k+1} = B^k + \frac{\gamma^k \gamma^{k\,T}}{\gamma^{k\,T} \delta^k} - \frac{B^k \delta^k \delta^{k\,T} B^{k\,T}}{\delta^{k\,T} B^k \delta^k} \tag{20}$$

with γ^k defined by

$$\gamma^k = \nabla L\left(\lambda^{k+1} Q^{k+1}, \mu^{k+1}, \eta^{k+1}\right) - \nabla L\left(\lambda^k, Q^k, \mu^{k+1}, \eta^{k+1}\right). \tag{21}$$

Second, to have global convergence performance, a line search method [33] is used to replace the full Newton step $(\lambda^{k+1}, Q^{k+1})^T = (\lambda^k, Q^k)^T + \delta^k$ by $(\lambda^{k+1}, Q^{k+1})^T =$

Algorithm 1 d-P-R-D Optimization-Based Determination of the Optimal Pair of Search Range λ^* and QP QP*

Initialization Step

Set an initial primal/dual point $(\lambda^0, Q^0, \mu^0, \eta^0)$ with $\mu^0 \geq 0$ and $\eta^0 \geq 0$, and a positive definite matrix B^0.

Let $k = 0$, and go to the iteration step.

Iteration Step

At kth iteration:

1. Solve the quadratic subproblem (19), with $\nabla^2 L(\lambda^k, Q^k, \mu^k, \eta^k)$ replaced by B^k, to obtain δ^k together with a set of Lagrange multipliers (μ^{k+1}, η^{k+1}). If μ^{k+1} or η^{k+1} is negative, then project it onto the set of nonnegative real numbers.
2. If $\delta^k = 0$, which shows that $(\lambda^k, Q^k, \mu^{k+1}, \eta^{k+1})$ satisfies the KKT conditions of problem (16), or $k+1$ exceeds the predefined maximum number of iterations, then map quantization step size Q^k to the corresponding QP QP^k, and go to the decision step.
3. Find $(\lambda^{k+1}, Q^{k+1})^T = (\lambda^k, Q^k)^T + \alpha \cdot \delta^k$ according to (22) and (23).
4. Update B^k to a positive definite matrix B^{k+1} based on (20).
5. Set $k = k+1$, and return to step 1.

Rounding-off Decision Step

For the pair of (λ^k, QP^k) obtained after the iteration step, find two consecutive integer values of search range $\bar{\lambda}$ and $\bar{\lambda}+1$, and two consecutive QPs $\overline{\text{QP}}$ and $\overline{\text{QP}}+1$, such that $\bar{\lambda} \leq \lambda^k < \bar{\lambda}+1$ and $\overline{\text{QP}} \leq \text{QP}^k < \overline{\text{QP}+1}$.

From the above four possible combinations of search range ($\bar{\lambda}$ or $\bar{\lambda}+1$) and QPs ($\overline{\text{QP}}$ or $\overline{\text{QP}}+1$), select the one both achieving the minimum distortion and satisfying (16b)–(16d) as the optimal pair (λ^*, QP^*).

$(\lambda^k, Q^k)^T + \alpha \cdot \delta^k$, which defines the l_1 merit function as

$$l_1(\lambda, Q) = D(\lambda, Q) + \theta \cdot [\max\{0, R(\lambda, Q) - R_{\max}\} + \max\{0, d(\lambda, Q) - d_{\max}\}] \quad (22)$$

where θ is a positive penalty parameter and the step size α is chosen such that the l_1 merit function is reduced

$$l_1\left((\lambda^k, Q^k)^T + \alpha \cdot \delta^k\right) < l_1\left((\lambda^k, Q^k)^T + \delta^k\right). \quad (23)$$

Considering both of the above modifications, an improved SQP algorithm is proposed for the d-P-R-D optimization-based source RC problem, as illustrated by Algorithm 1. Note that the proposed d-P-R-D model can be extended to a different coding mode, such as IBPBP coding mode, where the proposed model is still applicable to P-frames. However, the d-P-R-D model of B-frames needs to be reformulated in a similar way by considering the different temporal prediction distances for different B-frames.

IV. Experimental Results

To evaluate the performance of the proposed d-P-R-D model and the d-P-R-D optimization-based source RC algorithm, we implement the proposed d-P-R-D model and associated source RC algorithm in JM18.2 [21] encoder, with the test video sequences *Bus* (QCIF), *Foreman* (CIF), *Mobile* (CIF), and *Coastguard* (CIF) at 30 frames/s, the IPPPP GOP structure, CABAC entropy coding, the maximum search range 16 of ME, the dynamic range 0–51 of QP, and one reference frame. The maximum power consumption level is measured in the percentage of the maximum power consumption P_0 of the video encoder.

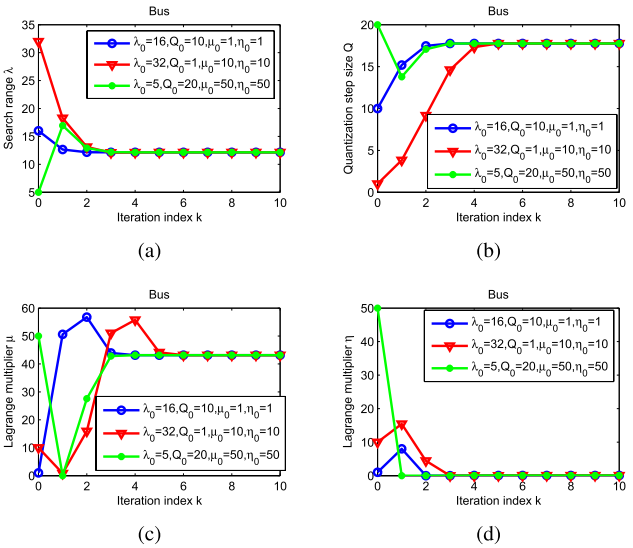


Fig. 9. Convergence behavior of (a) search range λ, (b) quantization step size Q, and Lagrange multipliers (c) μ, and (d) η for the *Bus* sequence, where $\theta = 1$, $R_{\max} = 1$ bits per pixel, $d_{\max} = 2.5$ s, and $P_{\max} = P_0$ is the maximum power consumption level of the video encoder, with three different sets of initial values.

A. Convergence Behavior and Model Accuracy

Fig. 9 shows the convergence behavior of the proposed d-P-R-D optimization-based source RC algorithm for the first ten frames of a video sequence. Here, the maximum power consumption level is set to 100%, i.e., $P_{\max} = P_0$. It can be observed that by the proposed RC algorithm, both primal (λ, Q) and dual variables (μ, η) can simultaneously and quickly converge to the corresponding optimal values in a few iterations. If the initial values of λ and Q are closer to the optimal solution, fewer number of iterations are needed for convergence. Within each iteration, on the other hand, it is only required to solve a QP optimization problem. In comparison, with the bisection search algorithm [15], both the feasible region of dual variables μ and η need an iterative bisection search, which greatly increases the number of iterations. Within each iteration of the bisection search algorithm, the entire feasible sets of primal variables λ and Q are exhaustively searched to find the optimal solution, which means that the duration of one iteration is much longer than that of the proposed RC algorithm.

In Fig. 9, the maximum source bit rate and encoding delay for the *Bus* video sequence are set to 1 bits per pixel and 2.5 s, respectively. After the rounding-off decision of the proposed RC algorithm, the minimum achievable distortion is 30.83 with the optimal parameters $\lambda^* = 12$ and $\text{QP}^* = 29$ ($Q^* = 18$). As validation, when the feasible sets of search range and QP are exhaustively searched for the first ten frames, the minimum distortion 31.35 is achieved with optimal parameters $\lambda^* = 11$ and $\text{QP}^* = 29$ ($Q^* = 18$). Therefore, the proposed RC algorithm can achieve the near-optimal performance in practice.

Fig. 10 shows the true 3-D Pareto surface and the estimated 3-D Pareto surface by the proposed RC algorithm of source distortion D, source rate R, and encoding time d, when the maximum power consumption level is set to 100%. A point

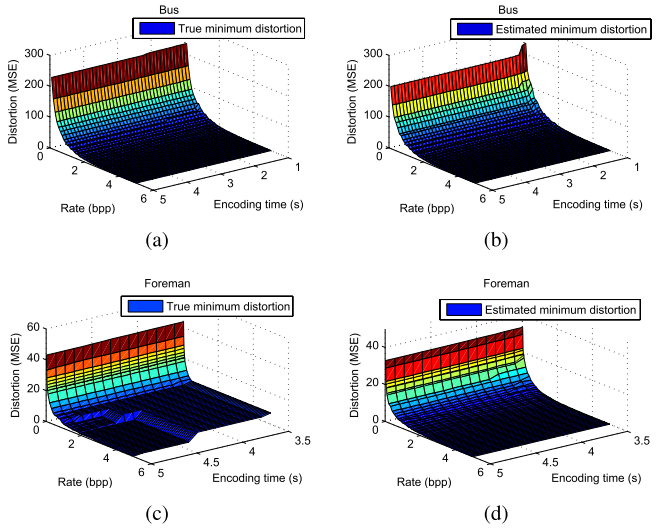


Fig. 10. 3-D Pareto surface of D, R, and d with $P_{\max} = P_0$ being the maximum power consumption level of the video encoder. (a) and (b) *Bus* sequence. (c) and (d) *Foreman* sequence.

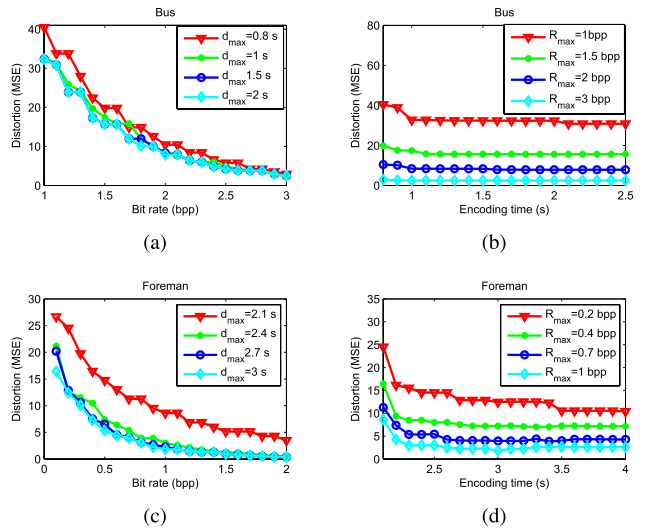


Fig. 11. D-R curves of the (a) *Bus* sequence and (c) *Foreman* sequence for different maximum encoding times $d_{\max}$, and D-d curves of the (b) *Bus* sequence and (d) *Foreman* sequence for different maximum bit rates $R_{\max}$, where $P_{\max} = P_0$ is the maximum power consumption level of the video encoder.

on the 3-D Pareto surfaces shows the minimum achievable distortion associated with given rate and encoding time constraints. It can also be observed that the model estimation of the proposed RC algorithm is quite accurate.

B. d-P-R-D Model Analysis

To view the proposed d-P-R-D model in more detail, Fig. 11 shows the D-R curves for different maximum encoding times, and D-d curves for different maximum source bit rates, when the maximum power consumption level is set to 100%. As the D-R curves in Fig. 11(a) and (c), for a given $d_{\max}$, $D_{\min}$ is a decreasing function of $R_{\max}$ and such curve becomes flat when $R_{\max}$ is relatively large, which corresponds to Shannon's source coding theory [35]. It is noted that the previous work on RC shows only one D-R curve of the similar shape, as shown in Fig. 11(a) and (c). This is because that within their

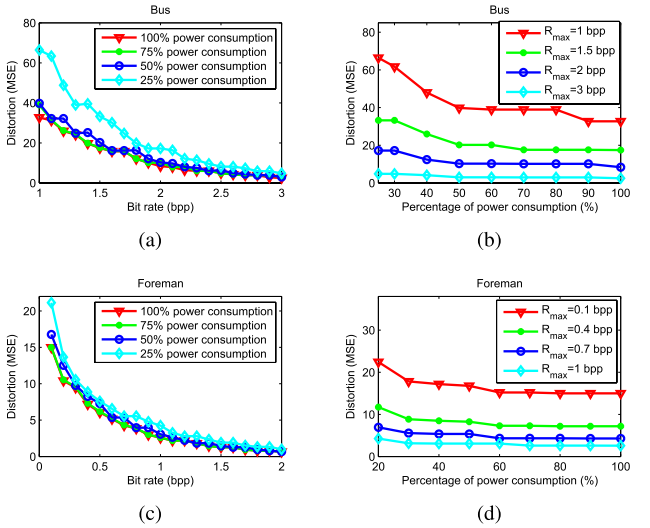


Fig. 12. D-R curves of the (a) *Bus* sequence and (c) *Foreman* sequence for different power consumption levels $P_{\max}$, and D-P curves of (b) *Bus* sequence and (d) *Foreman* sequence for different maximum bit rates $R_{\max}$, where $d_{\max}$ for the *Bus* sequence and the *Foreman* sequence are set to 1 and 4 s, respectively.

D-R models, the encoding time as well as encoding power are always assumed to be fixed but unspecified, which is a special case of the proposed d-P-R-D model. Hence, in this paper, the standard deviation σ of transformed residuals is fixed as a result of fixed λ and θ, and their RC is to tune QP since D and R are functions of QP alone. From the D-d curves in Fig. 11(b) and (d), it can be observed that $D_{\min}$ decreases with $d_{\max}$ but becomes quite flat for larger $d_{\max}$. This is because that ME has already achieved a global optimal motion vector at a certain value of search range, and hence continuously increasing the search range beyond that value contributes a little to the decrement of minimum achievable distortion.

Fig. 12 shows the D-R curves for different maximum power consumption levels, and D-P curves for different maximum source bit rates, with the maximum encoding delay fixed at 1 and 4 s for the *Bus* and *Foreman* sequences. It can be observed from Fig. 12(a) and (c) that for a given maximum power consumption level, the relationship of distortion and rate complies with Shannon's source coding theory. As shown by the D-P curves in Fig. 12(b) and (d), $D_{\min}$ decreases with the increment of the percentage of power consumption but becomes quite flat for larger power consumption level. The reason is same as the previous analysis, i.e., ME has already achieved a global optimal motion vector at a certain value of search range when the power consumption level is relatively large.

Fig. 13 shows the D-d curves for different maximum power consumption levels, and D-P curves for different maximum encoding times, with the maximum source bit rate fixed at 1.5 and 0.5 bits per pixel for the *Bus* and *Foreman* sequences. It can be observed that the similar analysis of the proposed d-P-R-D model is witnessed as well, i.e., $D_{\min}$ decreases with $d_{\max}$ but becomes quite flat for larger $d_{\max}$, and $D_{\min}$ decreases with the increment of the percentage of power consumption but becomes quite flat for larger power consumption level.

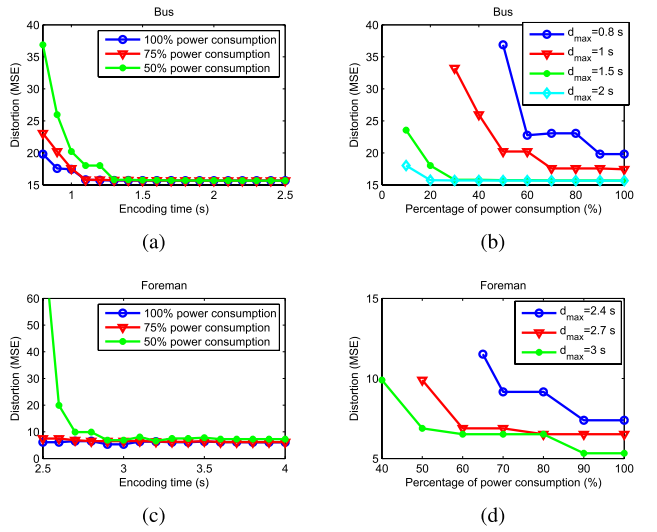


Fig. 13. D-d curves of the (a) *Bus* sequence and (c) *Foreman* sequence for different power consumption levels $P_{\max}$, and D-P curves of the (b) *Bus* sequence and (d) *Foreman* sequence for different maximum encoding times $d_{\max}$, where $R_{\max}$ for the *Bus* sequence and *Foreman* sequence are set to 1.5 and 0.5 bits per pixel, respectively.

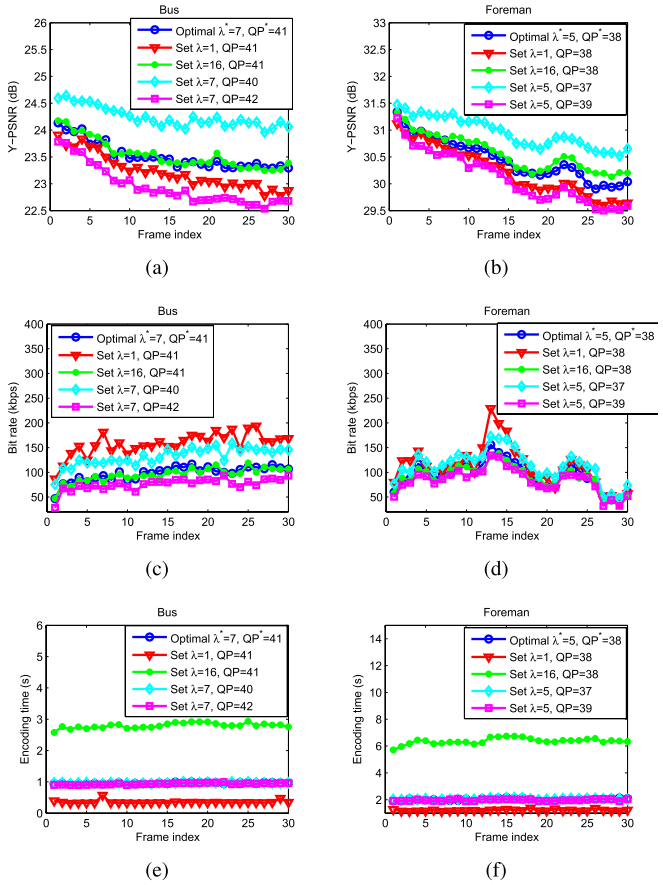


Fig. 14. (a) and (b) Objective quality, (c) and (d) bit rate, and (e) and (f) encoding time with different parameter pairs of (λ, QP) for the first 30 P-frames of the *Bus* and *Foreman* sequences, where $R_{\max} = 100$ kb/s, $P_{\max} = P_0$ is the maximum power consumption level of the video encoder, and $d_{\max}$ is set to 1 and 2 s for the *Bus* and *Foreman* sequences, respectively.

C. Performance Comparison

To demonstrate the performance of the proposed RC algorithm, the d-P-R-D model is derived from the

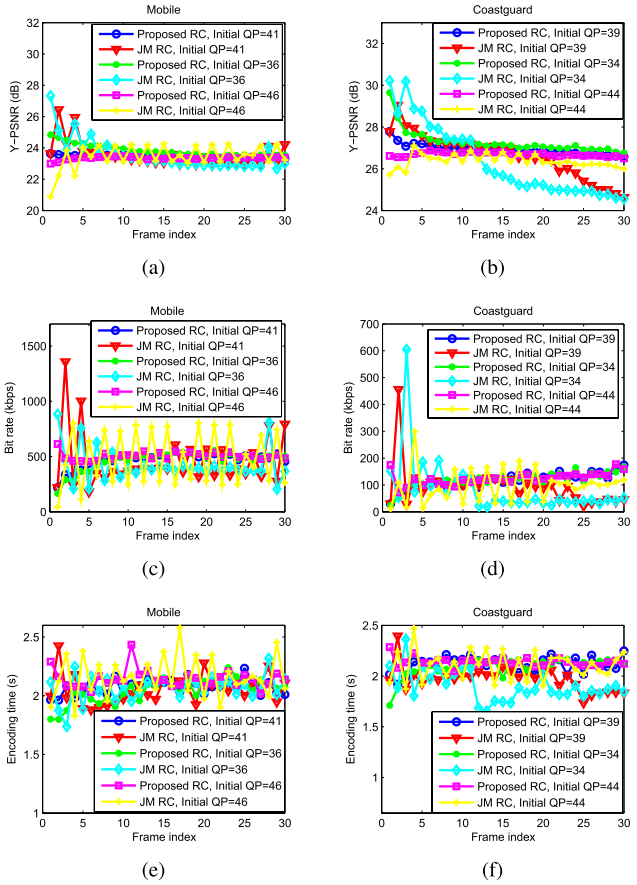


Fig. 15. (a) and (b) Objective quality, (c) and (d) bit rate, and (e) and (f) encoding time achieved by the proposed RC algorithm and the RC algorithm of JM with different setup of initial QP for the first I-frame, for the first 30 P-frames of the *Mobile* and *Coastguard* sequences, where search range is set to the optimal search range solved by the proposed algorithm, $d_{\max} = 2$ s, $P_{\max} = P_0$ is the maximum power consumption level of the video encoder, and $R_{\max}$ is set to 500 kb/s and 100 kb/s for the *Mobile* and *Coastguard* sequences, respectively.

first ten frames of a video sequence and the proposed RC algorithm acts on the entire video sequence. Note that when a new video scene is acquired, a new model $\sigma(\lambda, Q) = ae^{-b} + c + dQ$ needs to be fitted with three fitting parameters a, b, and c. To reduce such complexity in practice, since the function form of $\sigma(\lambda, Q)$ is already known and only three fitting parameters are unknown, we could choose a much smaller subset of empirical values with only a few configurations of λ and Q as training set and obtained the standard deviation model with the tradeoff of fitting accuracy. On the other hand, in the proposed RC algorithm, we only need to use the first several P-frames (e.g., ten frames in the experiment) to obtain the standard deviation model and to apply it as an estimated standard deviation model for the whole video sequence (e.g., 300 frames).

Therefore, the additional computational complexity introduced by the proposed RC algorithm per frame is not significant and thus can be neglected.

Fig. 14 shows the frame-wise objective quality in luminance-component PSNR (Y-PSNR), bit rate, and encoding time with different parameter pairs of (λ, QP) for the first 30 P-frames of the *Bus* and *Foreman* sequences. For the *Bus*

TABLE I

COMPARISON OF AVERAGE Y-PSNR, BIT RATE, AND ENCODING TIME (ET) UNDER DIFFERENT CONSTRAINTS OF MAXIMUM POWER CONSUMPTION LEVELS, BIT RATE, ENCODING DELAY, AND DIFFERENT SETUP IN THE INITIAL QP FOR THE FIRST I-FRAME, FOR THE FIRST 150 P-FRAMES OF *Foreman* AND *Coastguard* SEQUENCES

Sequence	P_{max} (%)		100				50			
	d_{max} (s)		2		3		2		3	
	R_{max} (kbps)		100	500	100	500	100	500	100	500
Foreman	Proposed RC QP^*	Y-PSNR(dB)	30.28	35.98	30.41	36.01	29.65	35.92	30.31	36.00
		Rate(kbps)	105.70	543.85	103.68	538.24	96.79	564.10	104.77	542.50
		ET(s)	2.14	2.13	3.12	2.85	2.03	1.98	3.11	3.08
	JM RC QP^*	Y-PSNR(dB)	30.04	35.67	30.27	35.73	29.68	35.49	30.17	35.71
		Rate(kbps)	100.31	494.69	102.02	494.46	99.10	494.40	100.92	494.73
		ET(s)	2.06	2.12	3.06	2.87	2.05	1.88	3.07	3.12
	Proposed RC $QP^* - 10$	Y-PSNR(dB)	30.72	36.09	30.81	36.13	30.08	36.05	30.74	36.12
		Rate(kbps)	106.56	544.56	104.01	539.72	96.79	563.47	105.00	541.28
		ET(s)	2.10	2.20	3.03	2.82	1.94	1.88	2.96	3.17
	JM RC $QP^* - 10$	Y-PSNR(dB)	30.14	35.50	30.30	35.54	29.82	35.26	30.27	35.52
		Rate(kbps)	96.85	470.75	95.43	472.71	95.38	471.16	97.16	472.58
		ET(s)	2.00	2.13	2.96	2.82	2.03	1.97	2.89	3.06
	Proposed RC $QP^* + 10$	Y-PSNR(dB)	30.02	35.94	30.14	35.96	29.35	35.89	30.05	35.95
		Rate(kbps)	109.09	550.55	107.88	546.15	99.91	569.65	108.34	549.28
		ET(s)	2.13	2.10	3.15	2.84	2.01	2.02	3.24	3.08
	JM RC $QP^* + 10$	Y-PSNR(dB)	29.83	35.61	29.93	35.67	29.45	35.42	29.85	35.64
		Rate(kbps)	103.49	502.03	101.69	501.55	102.84	501.23	101.52	502.01
		ET(s)	2.10	2.07	3.05	2.86	1.94	1.99	3.02	3.15
Coastguard	Proposed RC QP^*	Y-PSNR(dB)	24.88	29.09	24.97	29.15	24.82	29.04	24.93	29.11
		Rate(kbps)	90.34	515.00	89.89	512.65	93.18	520.20	89.75	513.51
		ET(s)	1.98	1.86	2.87	3.27	2.14	1.95	3.07	2.77
	JM RC QP^*	Y-PSNR(dB)	24.96	28.93	25.08	29.07	24.88	28.84	25.02	28.98
		Rate(kbps)	98.24	494.75	99.81	496.15	99.85	495.73	99.30	496.02
		ET(s)	1.93	1.89	2.78	3.32	2.13	1.98	3.09	2.91
	Proposed RC $QP^* - 10$	Y-PSNR(dB)	25.06	29.19	25.17	29.26	24.99	29.13	25.13	29.21
		Rate(kbps)	93.22	516.94	92.31	512.48	94.96	523.92	91.52	514.32
		ET(s)	2.07	1.99	2.79	3.25	2.17	1.90	3.01	3.07
	JM RC $QP^* - 10$	Y-PSNR(dB)	24.82	28.73	24.94	28.85	24.72	28.63	24.91	28.75
		Rate(kbps)	90.31	473.18	89.51	473.87	89.64	473.19	89.67	474.53
		ET(s)	2.01	1.84	2.73	3.19	2.19	1.97	2.99	2.89
	Proposed RC $QP^* + 10$	Y-PSNR(dB)	24.83	29.06	24.89	29.12	24.75	29.02	24.85	29.07
		Rate(kbps)	94.32	522.93	93.12	515.92	97.79	524.26	94.98	520.45
		ET(s)	2.13	2.01	2.85	3.16	2.27	2.04	3.04	2.90
	JM RC $QP^* + 10$	Y-PSNR(dB)	24.63	28.90	24.70	29.00	24.55	28.82	24.69	28.92
		Rate(kbps)	100.29	499.67	100.76	502.45	100.48	500.01	100.96	504.53
		ET(s)	2.07	1.92	2.69	3.14	2.13	1.97	3.02	2.91

sequence, we set $R_{\max} = 100$ kbps, $d_{\max} = 1$ s, and $P_{\max} = P_0$. The optimal parameter pair determined by the proposed RC algorithm is $(\lambda^*, \text{QP}^*) = (7, 41)$. It can be observed that when we vary the value of QP and fix the search range at the optimal value of seven, even though the overall quality [e.g., setting $(\lambda, \text{QP}) = (7, 40)$] may be higher than the optimal Y-PSNR of the proposed RC algorithm, the bit rate constraint has been violated, which shows that the corresponding solution is not feasible. When QP is fixed at the optimal value of 41, on the other hand, setting the search range greater than the optimal value of seven does not contribute much to improve the coding efficiency, while any search range less than the optimal value would lead to performance degradation. For example, when we set λ to the maximum value of 16, the Y-PSNR and bit rate performance is similar to the optimal solution, while the encoding time is much greater than the encoding time constraint. If $\lambda = 1$, the Y-PSNR is lower than the optimal solution while the bit rate violates the bit rate constraint. Therefore, the proposed RC algorithm can find the search range threshold over which the quality does not make much improvement, which is particularly useful in saving encoding time and computation power. That is, since the source coding complexity mainly depends on the ME complexity, if we know the coding efficiency at a lower cost (smaller search range) is similar to at a higher cost (larger search range), it is unnecessary to spend the valuable computational resources on extra ME. Thus, the power consumption and computation time can be saved. The similar observation can also be found for the *Foreman* sequence, where $R_{\max} = 100$ kbps, $d_{\max} = 2$ s, and $P_{\max} = P_0$, and the optimal parameter pair determined by the proposed RC algorithm is $(\lambda^*, \text{QP}^*) = (5, 38)$.

In Fig. 15, we compare the frame-wise performance of the proposed RC algorithm with the RC scheme in JM18.2 over the other test sequences. Note that since the JM RC scheme fails to deal with the delay and power constraints, to meet the maximum delay and power constraints and thus have a fair comparison with the proposed RC scheme, here, we set the search range for the JM RC scheme as the optimum value obtained by the proposed RC scheme.

It can be observed that, however, the performance of the JM RC scheme depends greatly on the initial QP value of the first I-frame. According to [11], the selection of the initial quantization step size for the first I-frame is very critical for model-based RC algorithm, since the R-D of the first frame

can affect the coding efficiency of the subsequent frames. Therefore, for JM RC, if the initial quantization step size is set to be too small, the bit rate for the first few frames would be much larger than the target bit rate. To make the average bit rate meet the target bit rate, the bit rates allocated for the rest few frames would be much smaller by increasing quantization step sizes for these frames. For example, for different setups of initial QP for the first I-frame, the results of Y-PSNR, bit rate, and encoding time per frame obtained by the JM RC scheme are significantly different, while the results of the proposed RC algorithm are relatively similar and not affected too much by different initial QP values. In addition, the JM RC scheme fails to get a stable performance for the entire sequence, which is required for a better user experience. The corresponding Y-PSNR, bit rate, and encoding time versus frame index curves are quite fluctuated and unstable. Although the Y-PSNR values for some frames might be higher than those of the proposed RC algorithm, the bit rate and encoding time constraints are violated. In contrast, the selection of the initial quantization step size for the first I-frame would have less impact on the performance of the proposed RC algorithm. The proposed RC algorithm can achieve smooth and stable Y-PSNR performance while the other constraints are satisfied.

The reason for the stability of the proposed RC algorithm is as follows. For each video sequence, specifically, the proposed d-P-R-D model is derived based on the first several frames, and thus can characterize more accurately the statistics in that video sequence. When adopting such d-P-R-D model in the proposed model-based RC algorithm, the optimal coding parameters can be determined in accordance with different video sequences. In addition, for each sequence, λ^* and Q^* would remain the same for all frames, which makes the actual bit rate and Y-PSNR more stable among different frames. For JM RC scheme, the quantization step size is tuned and varied for each frame in accordance with general distortion-quantization and rate-quantization models, which are independent of different video contents. Therefore, the optimal quantization step size determined by JM RC scheme would change frame by frame to meet the target bit rate while minimizing the coding distortion. In addition, the actual bit rate and Y-PSNR would fluctuate with such change in quantization step size among different frames.

Table I shows the comparison of average Y-PSNR, bit rate, and encoding time for the first 150 P-frames of *Foreman* and *Coastguard* sequences, respectively. The maximum achieved Y-PSNR results can verify the analysis of the d-P-R-D model in Section IV-B. It can also be observed that the selection of the initial quantization step size for the first I-frame would have less impact on the average Y-PSNR of the proposed RC scheme. That is, for the proposed RC scheme, the average Y-PSNR performance under different selection of the initial quantization step size is stable and usually better than that of the JM RC scheme.

V. Conclusion

In this paper, we derived the analytical d-P-R-D model for IPPPP coding mode in H.264/AVC to investigate the relationship among video encoding time, power, rate, and distortion. On the basis of the proposed d-P-R-D model, a model-based source RC problem has been formulated to minimize the encoding distortion under the constraints of rate, delay, and power. To solve the RC problem, we proposed a practical algorithm to iteratively update the primal and dual variables using both the KKT conditions and the SQP method. The experimental results have verified the accuracy of the proposed d-P-R-D model and demonstrated the optimization performance of the model-based source RC algorithm. The d-P-R-D model and the model-based RC provided a theoretical basis and a practical guideline for the cross-layer system design and performance optimization in wireless video communication under delay and energy constraints. To further tackle the issue of bandwidth fluctuations and higher packet losses in wireless transmissions, our future work will focus on applying the proposed d-P-R-D approach to joint resource allocation and control for the entire wireless video communication system, which aims at minimizing the end-to-end distortion under the constraints of the transmission bandwidth, the end-to-end delay, and the total available power supply of the wireless video communication system.

References

[1] Z. He, Y. Liang, L. Chen, I. Ahmad, and D. Wu, "Power-rate-distortion analysis for wireless video communication under energy constraints," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 5, pp. 645–658, May 2005.

[2] (2013). *Cisco Visual Networking Index (VNI) Forecast* [Online]. Available: http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.pdf

[3] X. Li, N. Oertel, A. Hutter, and A. Kaup, "Laplace distribution based Lagrangian rate distortion optimization for hybrid video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 2, pp. 193–205, Feb. 2009.

[4] C. Y. Hsu, A. Ortega, and M. Khansari, "Rate control for robust video transmission over burst-error wireless channels," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 5, pp. 756–773, May 1999.

[5] Q. Chen, "Image and video processing for denoising, coding, and content protection," Ph.D. dissertation, Dept. Electr. Comput. Eng., Univ. Florida, Gainesville, FL, USA, 2011.

[6] S. Soltani, K. Misra, and H. Radha, "Delay constraint error control protocol for real-time video communication," *IEEE Trans. Multimedia*, vol. 11, no. 4, pp. 742–751, Jun. 2009.

[7] Q. Chen and D. Wu, "Delay-rate-distortion model for REA-time video communication," to be published.

[8] S. Khan, Y. Peng, E. Steinbach, M. Sgroi, and W. Kellerer, "Application-driven cross-layer optimization for video streaming over wireless networks," *IEEE Commun. Mag.*, vol. 44, no. 1, pp. 122–130, Jan. 2006.

[9] Z. Li, F. Pan, K. Lim, G. Feng, X. Lin, and S. Rahardja, "Adaptive basic unit layer rate control for JVT," presented at the 7th JVT Meeting, Pattaya, Thailand, Mar. 2003.

[10] D. Kwon, M. Shen, and C. 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.

[11] H. Wang and S. Kwong, "Rate-distortion optimization of rate control for H.264 with adaptive initial quantization parameter determination," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 1, pp. 140–144, Jan. 2008.

[12] B. Yan and K. Sun, "Joint complexity estimation of I-frame and P-frame for H.264/AVC rate control," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 5, pp. 790–798, May 2012.

[13] S. Ma, W. Gao, and Y. Lu, "Rate-distortion analysis for H.264/AVC video coding and its application to rate control," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 12, pp. 1533–1544, Dec. 2005.

[14] I. E. G. Richardson. (2003). *H.264/MPEG-4 Part 10: Transform and Quantization* [Online]. Available: http://www.vcodex.com

[15] Z. Chen and D. Wu, "Rate-distortion optimized cross-layer rate control in wireless video communication," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 3, pp. 352–365, Mar. 2012.
[16] A. Aminlou, Z. NajafiHaghi, M. Namaki-Shoushtari, and M. R. Hashemi, "Rate-distortion-complexity optimization for VLSI implementation of integer motion estimation in H.264/AVC encoder," in *Proc. IEEE ICME*, Jul. 2011, pp. 1–6.
[17] J. Vanne, M. Viitanen, T. Hamalainen, and A. Hallpuro, "Comparative rate-distortion-complexity analysis of HEVC and AVC video codecs," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1885–1898, Dec. 2012.
[18] G. Corrêa, P. Assuncao, L. A. da Silva Cruz, and L. Agostini, "Adaptive coding tree for complexity control of high efficiency video encoders," in *Proc. IEEE PCS*, May 2012, pp. 425–428.
[19] H. M. Hang and J. J. Chen, "Source model for transform video coder and its application. I. Fundamental theory," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 2, pp. 287–298, Apr. 1997.
[20] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.
[21] K. Sühring. (2012, Feb.). *H.264/AVC Reference Software JM18.2* [Online]. Available: http://iphome.hhi.de/suehring/tml/download/
[22] C. Li, D. Wu, and H. Xiong, "Delay-power-rate-distortion model for wireless video communication under delay and energy constraints," Dept. Electr. Comput. Eng., Univ. Florida, Gainesville, FL, USA, Tech. Rep., 2014 [Online]. Available: http://www.wu.ece.ufl.edu/mypapers/dPRD-TR.pdf.
[23] R. Min, T. Furrer, and A. Chandrakasan, "Dynamic voltage scaling techniques for distributed microsensor networks," in *Proc. IEEE Comput. Soc. Workshop VLSI*, Apr. 2000, pp. 43–46.
[24] J. R. Lorch and A. J. Smith, "Improving dynamic voltage scaling algorithms with PACE," *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 29, no. 1, pp. 50–61, Jun. 2001.
[25] T. D. Burd and R. W. Brodersen, "Processor design for portable systems," *J. VLSI Signal Process.*, vol. 13, no. 2, pp. 203–221, 1996.
[26] E. Lam and J. Goodman, "A mathematical analysis of the DCT coefficient distributions for images," *IEEE Trans. Image Process.*, vol. 9, no. 10, pp. 1661–1666, Oct. 2000.
[27] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[28] H. Everett, "Generalized Lagrange multiplier method for solving problems of optimum allocation of resources," *Oper. Res.*, vol. 11, no. 3, pp. 399–417, Jun. 1963.
[29] X. Li, N. Oertel, A. Hutter, and A. Kaup, "Advanced Lagrange multiplier selection for hybrid video coding," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2007, pp. 364–367.
[30] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers [speech coding]," *IEEE Trans. Acoust., Speech Signal Process.*, vol. 36, no. 9, pp. 1445–1453, Sep. 1988.
[31] A. Ortega and K. Ramchandran, "Rate-distortion methods for image and video compression," *IEEE Signal Process. Mag.*, vol. 15, no. 6, pp. 23–50, Nov. 1998.
[32] D. Bertsekas, *Nonlinear Programming*. Belmont, MA, USA: Athena Scientific, 1999.
[33] B. Chachuat, *Nonlinear and Dynamic Optimization: From Theory to Practice*. Zürich, Switzerland: Automatic Control Lab., 2007.
[34] J. Nocedal and S. J. Wright, *Numerical Optimization*. New York, NY, USA: Springer-Verlag, 2006.
[35] T. Cover and J. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 2006.

Chenglin Li (S'13) received the B.S. and M.S. degrees in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2007 and 2009, respectively, where he is currently working toward the Ph.D. degree.

He was with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA, from 2011 to 2013 as a Visiting Ph.D. Student. His research interests include network-oriented image/video processing and communication, and network-based optimization for video sources.

Dapeng Wu (S'98–M'04–SM'06–F'13) received the Ph.D. degree in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, USA, in 2003.

He has been with the faculty of Electrical and Computer Engineering Department, University of Florida, Gainesville, FL, USA, since 2003, where he is a Professor. His research interests include networking, communications, signal processing, computer vision, and machine learning.

Dr. Wu received the University of Florida Research Foundation Professorship Award in 2009, AFOSR Young Investigator Program (YIP) Award in 2009, ONR YIP Award in 2008, NSF CAREER Award in 2007, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY Best Paper Award in 2001, and the Best Paper Awards in IEEE GLOBECOM in 2011 and the International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks (QShine) in 2006.

Hongkai Xiong (M'01–SM'10) received the Ph.D. degree in communication and information system from Shanghai Jiao Tong University (SJTU), Shanghai, China, in 2003.

He is with the Department of Electrical Engineering, SJTU, where he is currently a Professor. From 2007 to 2008 he was with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, USA, as a Research Scholar. He has published more than 70 international journal/conference papers. His research interests include source coding/network information theory, signal processing, computer vision and graphics, and statistical machine learning.

Dr. Xiong has been involved with various IEEE Conferences as a Technical Program Committee Member. He is a member of the Technical Committee on Signal Processing of Shanghai Institute of Electronics. He received the New Century Excellent Talents in University Award in 2009 and the Young Scholar Award of SJTU in 2008. In SJTU, he directs Image, Video, and Multimedia Communications Laboratory and Multimedia Communication area in the Key Laboratory of Ministry of Education of China, Intelligent Computing and Intelligent System, which is also co-granted by Microsoft Research.