

# Video Texture and Motion Based Modeling of Rate Variability-Distortion (VD) Curves

Geert Van der Auwera, Martin Reisslein, and Lina J. Karam

**Abstract**—We examine and model the bit rate variability-distortion (VD) curve of -4 Part 2 variable bit rate (VBR) video encodings. The VD curve has important applications for evaluating the statistical multiplexing of streaming video. We show that the concave VD curve shape at high compression ratios, or equivalently large quantization scales, is influenced by both the texture and the motion information. Based on this insight, we first develop a general VD curve model by analytically expressing the VD curve in terms of elementary statistics (mean, variance, covariance) of the numbers of motion and texture coding bits. In a second step we develop and validate linear and quadratic models for the elementary texture and motion bit statistics, whereby the model parameters are obtained from only two sample encodings. The texture and motion bit models are then employed in our general VD curve model. This work extends our previous work on piecewise models of the VD curve. The texture and motion based VD model obtained from two sample encodings has comparable accuracy to a piecewise model based on three sample encodings. In addition, the texture and motion based VD model provides fundamental insights into how texture and motion affect traffic variability.

**Index Terms**—Communication systems, -4 Part 2, rate variability-distortion, statistical multiplexing, variable bit rate, video coding, video content, video quality, video streaming, video traffic.

## I. INTRODUCTION

A RECENT study [1] has documented the concave shape of the rate variability-distortion (VD) curve of open-loop or variable bit rate (VBR) -4 Part 2 [2] encoded video and proposed a crude model for the VD curve. The VD curve is the coefficient of variation (standard deviation normalized by the mean) of the frame size (in bits) as a function of the quantization scale or distortion metric. In [1], the VD curve has been approximated by a piecewise model. The model is based on up to four encoding samples corresponding to four different quantization scales. In addition, this piecewise model does not provide insights into the properties of the encoded video frames that result in the concave VD curve shape.

In this study, we build on [1] by examining the underlying effects leading to the concave VD curve and by developing and validating a significantly refined VD curve model for encoded video. More specifically, we first express the VD curve of I-,

P-, and B-frames in terms of elementary statistics (mean, variance, covariance) of the numbers of bits required to code the texture and motion information in the frames. These individual VD frame models are then combined in a general VD model for a video sequence. In a second step, we develop and validate quadratic models for the mean and variance of the texture information (bits) in a frame as a function of the quantization scale. We also develop linear models for the mean and variance of the motion information (bits) in a frame, thus extending [3] where a quadratic model was used for the rate-distortion function of the entire frame size. These linear and quadratic models of the texture and motion bits are then employed within our general VD curve model.

In the course of this modeling work, we demonstrate that both the texture and motion bits make significant contributions to the overall concave shape of the VD curve. We find that given the frame size contributions for encodings with only two different quantization scales, this novel model accurately predicts the VD curve across a wide range of quantization scales. This is a significant improvement over the crude model in [1] where encodings for three or more different quantization scales were required. The refined VD curve model proposed in this paper not only provides fundamental insights into how texture and motion in an encoded video contribute to its traffic variability, but also provides a practical method for estimating the VD curve from only two sample encodings. In addition, the general VD curve model derived in this paper is independent from the linear and quadratic texture and motion bit models. Other models for the texture and motion bit statistics can be employed within our general VD curve model, making it a useful basis for future research.

The VD curve has important applications in evaluating statistical multiplexing, which is often employed in video transport over networks. For instance, the VD curve is useful for admission control, i.e., for determining whether a set of video streams can be transmitted over a link subject to a prescribed loss probability limit. The VD curve model developed in this paper can be used to accurately determine the frame size variance for a range of quantization scales from only two sample encodings. This VD curve model in turn can be used in conjunction with the admission control methodology outlined in [1] to assess the loss probability for a range of quantization scales. Similarly, the VD curve model can be employed to assess the utility (revenue) earned from a streaming video service across a range of video quality levels.

This paper is organized as follows. In Section I-A, we discuss related work and position our research on the VD curve. Metric definitions and experimental setup are covered in Section II. The VD curve modeling principle is introduced in Section III-A. In

Manuscript received June 23, 2006; revised December 23, 2006. This work was supported in part by the US National Science Foundation Career Grants ANI-0133252 and ANI-0136774. Preliminary results from this study appear in "Video Texture and Motion Based Modeling of Rate Variability-Distortion (VD) Curves of I-, P-, and B-Frames," *Proc. of IEEE Int. Conf. on Multimedia & Expo*, pages 1405–1408, Toronto, Canada, July, 2006.

The authors are with the Department of Electrical Engineering, Arizona State University, Goldwater Center, AZ 85287-5706 USA (e-mail: geert.vanderauwera@asu.edu; karam@asu.edu; reisslein@asu.edu).

Digital Object Identifier 10.1109/TBC.2007.903624

Section III-B we first derive the VD curve models for the I-, P-, and B-frames. The combined general VD curve model for a full video sequence is presented in Section III-C. The quadratic and linear models for the texture and motion bit statistics of video frames are developed and validated in Sections IV-A and IV-B. In Section IV-C, the VD curve models for the I-, P-, and B-frames are validated, and in Section IV-D we evaluate the full VD model. The sensitivity of choosing the two quantization scales is analysed in Section V. We summarize our conclusions in Section VI.

#### A. Related Work

This study relates most closely to research on the rate-distortion (RD) characteristics of encoded video and research on the analysis and modeling of video traffic. The RD characteristics of encoded video give the relationship between the (mean) bit rate and the video quality (and encoder quantization scale), see e.g., [4]. RD modeling has been a prolific research area for many years and a wide variety of modeling approaches have been examined, ranging from models primarily based on the source video statistics, e.g., [5], to models primarily based on interpolating sample points of the RD curve, e.g., [6], as well as models incorporating analytical and empirical insights, e.g., [7], [8]. The RD models are important for rate control in video coding, e.g., [9], and the allocation of mean bit rates to video streams for network transport [10], [11]. Our study differs from the research on the RD characteristics in that we examine the relationship between the variability of the bit rate and the video quality (and quantization scale). That is, our focus is on the second order statistic of the video traffic, whereas the existing RD studies have focused on the first order statistic of the video traffic.

Research on the statistical analysis of video traffic and video traffic models has also received significant interest, see for instance [12]–[18]. These existing works focus on analysing the statistics of the traffic, which includes the bit rate variability, for a video encoding with a fixed quality level or quantization scale. In contrast, in this paper we examine the bit rate variability of the video as a function of the quality level/quantization scale. A detailed understanding of this bit rate variability-quality level relationship is important for the design and operation of video transport mechanisms, for individual streams, e.g., [19]–[21], for multiple streams of a fixed quality level, e.g., [22]–[24], as well as multiple streams with quality adaptation, see e.g., [25]–[28]. These transport mechanisms typically allocate buffer and bandwidth resources and employ dynamic scheduling strategies to accommodate the video traffic variability. Our study complements this literature by uncovering and characterizing the fundamental relationships between the rate variability produced by the open-loop encoder and the different quantization scales/quality levels.

## II. METRIC DEFINITIONS AND EVALUATION SETUP

The rate variability-distortion (VD) curve relates the bit rate variability of an encoded video sequence to its average picture quality. More specifically, the VD curve relates the coefficient of variation  $CoV$  of encoded frame sizes (number of bits) to the average picture quality (e.g., peak signal-to-noise ratio or

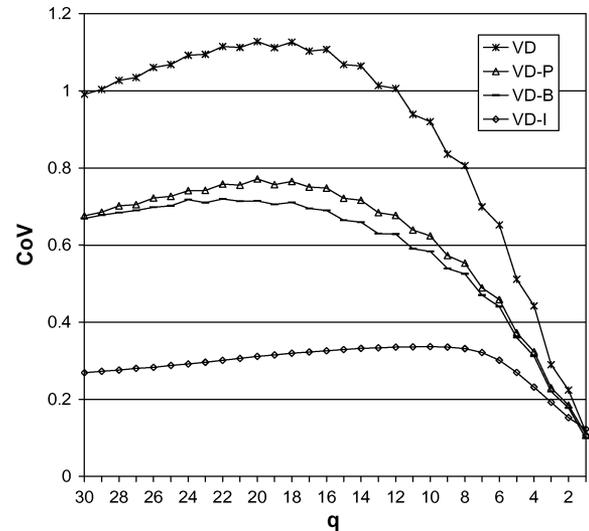


Fig. 1. Concave VD curves for the movie segment from *Star Wars V*, encoded with -4.

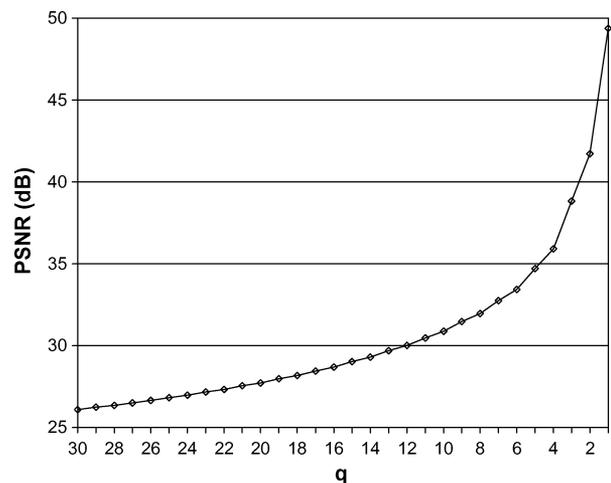


Fig. 2. Average picture quality of the movie segment from *Star Wars V* as a function of quantization scale  $q$ .

PSNR) or the quantization scale  $q$ . The  $CoV_q$  of the encoded frame sizes of a given video sequence for a particular quantization scale  $q$  is defined as

$$CoV_q = \frac{\sigma_q}{\bar{R}_q} \quad (1)$$

with  $\sigma_q$  denoting the standard deviation and  $\bar{R}_q$  the mean of the frame sizes (in bits) of the encoded video sequence. The VD curve has a characteristic concave or “hump” shape as depicted in Fig. 1, separately for I-, P-, B-frames, and combined for all frame types for a segment of the movie *Star Wars V: The Empire Strikes Back* (1000 QCIF frames). Due to the concave shape, there exists a quantization scale  $q$  at which the bit rate variability reaches a maximum. In Fig. 1, the abscissa contains the  $q$  scale and decreases from value 30 to 1. This corresponds to the increasing average picture quality as is illustrated in Fig. 2, where the average PSNR is plotted as a function of quantization

TABLE I  
OVERVIEW OF SCENES CLASSIFIED IN 5 MOTION CLASSES

Video	Scene #	Length (frames)	Motion Class	Description
<i>Football</i>	298	227	I	Intel logo with moving background (trailer).
<i>Star Wars IV</i>	274	443	I	Princess Leia's hologram pleas for help.
<i>Terminator</i>	384	520	I	Talk between two humans in a tunnel.
<i>Football</i>	299	367	II	NFL logo animation.
<i>Star Wars IV</i>	117	391	II	Slow zoom on R2D2 and C3PO marching in the desert.
<i>Terminator</i>	462	275	II	Terminator makes a call.
<i>Football</i>	557	266	III	Subway train GAP commercial.
<i>Star Wars IV</i>	115	112	III	Darth Vader gives orders and walks away.
<i>Terminator</i>	628	140	III	Gas truck makes U-turn.
<i>Football</i>	184	111	IV	Robot slides to camera (trailer).
<i>Star Wars IV</i>	165	89	IV	R2D2 is captured by sandpeople.
<i>Terminator</i>	262	69	IV	Car breaks through entrance doors.
<i>Football</i>	336	86	V	Camera follows running player.
<i>Star Wars IV</i>	632	46	V	Fight in a bar on Tatooine.
<i>Terminator</i>	441	253	V	A picture burns.

scale  $q$ . The PSNR is an objective measure of the quality of a reconstructed video frame  $R(x, y)$  with respect to the uncompressed frame  $F(x, y)$  and for an  $N_x \times N_y$  frame consisting of 8-bit pixel values is computed as:

$$MSE = \frac{1}{N_x \cdot N_y} \sum_{x=0}^{N_x-1} \sum_{y=0}^{N_y-1} [F(x, y) - R(x, y)]^2, \quad (2)$$

$$PSNR = 10 \cdot \log \frac{255^2}{MSE}. \quad (3)$$

Other objective video quality metrics exist that correlate better with the quality perception of a human observer, e.g. VQM [29], [30]. These metrics can be applied to the reconstructed video sequences and a transformation curve between  $q$  and the metric values can be plotted, as is done for the PSNR in Fig. 2. Whatever quality metric is applied, it will have a monotonically increasing or decreasing trend from low quality for  $q = 30$  to high quality for  $q = 1$ . Hence, the concave curve shape of the VD curve will be preserved, although it can be stretched or compressed. The following analysis and conclusions will therefore also be valid for other quality metrics.

#### A. Video Scenes Used in Evaluations

In addition to the *Star Wars V* segment, we employ the same videos as in [1], namely scenes from two movies, *Star Wars IV*, *The Terminator*, and a *Football* game, which are summarized in Table I. All videos are in QCIF format (176×144 pixels). We used the publicly available scene detection software [31] for the determination of the scene boundaries. Following [32], the scenes have been classified according to the level of motion into five motion classes ranging from motion class I for a low level of motion, to motion class V for a high level of motion. For video encoding, we use the same video codec as in [1], namely the *Microsoft v2.3.0* reference implementation of -4 Part 2 [2] in the simple profile. The group-of-pictures (GoP) structure that is used in this work consists of twelve frames, namely one I-frame, three P-frames, and eight B-frames: *IBBPBBPBBPBB*.

### III. TEXTURE AND MOTION BASED VD CURVE MODEL

#### A. Texture vs. Motion Bits

In this paper, we apply a new modeling approach to the VD curve. We start from the insight that encoded video frames have as constituents texture, motion, and syntax bits. In case of the P- and B-frames, the texture bits represent the encoded prediction error information that remains after the often imperfect temporal prediction step in the video encoder. The motion bits represent side information (motion vectors) about the temporal prediction step and are required to reconstruct the P- and B-frames. In case of I-frames, the texture bits represent the actual encoded frame content. No motion bits are required to reconstruct the I-frames. The syntax bits assign a meaning to parts of the bit stream, e.g. to distinguish between texture and motion bits.

At low compression ratios, or equivalently, for small quantization scales, the motion and syntax bits are negligible compared to the bits required to encode the texture. With higher compression ratios, i.e., larger quantization scales, the texture information is significantly reduced. As a result, *for large quantization scales, the number of texture bits is comparable to the motion bits and, therefore, the motion information plays a significant role in the bit rate variability*. In other words, the concave VD curve shape at high compression ratios is influenced by the texture information but also by the motion information. We still consider the syntax bits negligible compared to the texture and motion bits, since they are more than an order of magnitude smaller.

This principle is illustrated in Fig. 3 where the average P- and B-frame sizes are depicted along with the average texture and motion bits for the *Star Wars V* segment. We observe that for  $q = 16$ , the P texture information is about twice the size of the motion information and decreases exponentially. For  $q = 24$ , the average number of P texture bits is even smaller than the number of P motion bits. For B-frames, this point is reached for  $q = 26$ . The average I-frame size curve is not shown in this figure, because it is more than three times larger than the average P-frame size curve. The I-frame sizes as a function of  $q$  have an exponential curve shape as well.

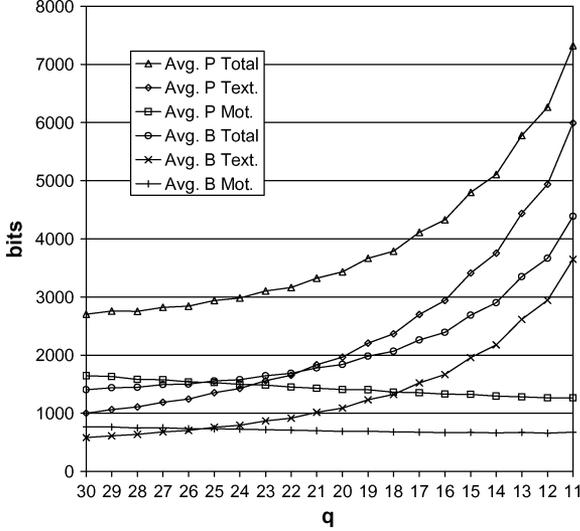


Fig. 3. Average P- and B-frame texture, motion bits, and total frame sizes as a function of quantization scale  $q$  for the *Star Wars V* video segment.

### B. VD Curve Model for I-, P-, and B-Frames

We denote  $X_{q,t,i}$  for the number of texture bits of frame  $i$  encoded with quantization scale  $q$ , and correspondingly  $X_{q,m,i}$  for the number of motion bits. For a sequence with  $M$  frames, we let  $\bar{R}_{q,t} = (1/M) \sum_{i=1}^M X_{q,t,i}$  and  $\sigma_{q,t}^2 = (1/(M-1)) \sum_{i=1}^M (X_{q,t,i} - \bar{R}_{q,t})^2$  respectively denote the (sample) mean and variance of the number of texture bits, and we define  $\bar{R}_{q,m}$  and  $\sigma_{q,m}^2$  analogously to denote the average and variance of the number of motion bits, respectively. We let  $cov_q(t, m)$  represent the (sample) covariance of the texture and motion information which is defined as  $cov_q(t, m) = (1/(M-1)) \sum_{i=1}^M (X_{q,t,i} - \bar{R}_{q,t}) \cdot (X_{q,m,i} - \bar{R}_{q,m})$ . We observe that the sum of the number of texture and motion bits approximately equals the total frame size, i.e., when we ignore the syntax bits:  $\bar{R}_q = \bar{R}_{q,t} + \bar{R}_{q,m}$ . With this approximation, we can express the  $CoV_q$  for the P- and B-frames as:

$$CoV_q^{(P,B)} = \frac{\sigma_q}{\bar{R}_q} = \frac{\sqrt{\sigma_{q,t}^2 + \sigma_{q,m}^2 + 2 \cdot cov_q(t, m)}}{\bar{R}_{q,t} + \bar{R}_{q,m}}, \quad (4)$$

which we refer to as the VD-P and VD-B models. For small  $q$  values, the motion bits are negligible compared to the texture bits and therefore (4) reduces to:

$$CoV_{small\ q}^{(P,B)} = CoV_q^{(I)} = \frac{\sigma_{q,t}}{\bar{R}_{q,t}}. \quad (5)$$

Equation (5) is applicable to the I-frames as well, since no motion information is present, and we refer therefore to (5) as the VD-I model.

### C. Combined VD Curve Model for Video Sequences

The next step in the modeling is combining the VD curve models for the I-, P-, and B-frames into a single VD curve model. The frame size variabilities expressed by the VD-I, VD-P, and VD-B models each contribute to the overall variability of the entire frame sequence. Intuition tells us that the relative contributions of the I-, P-, and B-frames to the overall frame size variability depend on the group-of-pictures (GoP) structure. The GoP structure that is used in this work consists of twelve frames and is given by: *IBBPBBPBBPBB*, i.e., one I-frame, three P-frames and eight B-frames. Denote the I-, P-, and B-frame fractions by:  $a^{(I)}$ ,  $a^{(P)}$ , and  $a^{(B)}$ . The fractions for our chosen GoP structure are:  $a^{(I)} = (1/12)$ ,  $a^{(P)} = (1/4)$ , and  $a^{(B)} = (2/3)$ . In the following, we assume there are an integer number  $N$  of GoPs in the sequence so that the total number of frames equals  $12 \times N$ . For a particular quantization scale  $q$ , the overall average of frame sizes  $\bar{R}_q$  can therefore be expressed as follows, with  $X_{q,i}$  denoting individual total frame sizes:

$$\begin{aligned} \bar{R}_q &= \frac{1}{12N} \left( \sum_{i=1}^N X_{q,i}^{(I)} + \sum_{i=1}^{3N} X_{q,i}^{(P)} + \sum_{i=1}^{8N} X_{q,i}^{(B)} \right) \quad (6) \\ &= \frac{1}{12} \cdot \bar{R}_q^{(I)} + \frac{1}{4} \cdot \bar{R}_q^{(P)} + \frac{2}{3} \cdot \bar{R}_q^{(B)}. \quad (7) \end{aligned}$$

Hence, the overall average of the frame sizes is the weighted sum of the averages per frame type (I-P-B) with weighting coefficients equal to the fractions of each frame type in the GoP structure. Furthermore, we have  $\bar{R}_q^{(I,P,B)} = \bar{R}_{q,t}^{(I,P,B)} + \bar{R}_{q,m}^{(I,P,B)}$ .

To calculate the overall variance, we follow a similar approach, as detailed in Appendix A, and obtain for the overall variance  $\sigma_q^2$  for a particular quantization scale  $q$ :

$$\begin{aligned} \sigma_q^2 &= \frac{1}{12} \cdot \sigma_q^{2(I)} + \frac{1}{4} \cdot \sigma_q^{2(P)} + \frac{2}{3} \cdot \sigma_q^{2(B)} + \frac{1}{12} \cdot \left( \bar{R}_q^{(I)} \right)^2 \\ &\quad + \frac{1}{4} \cdot \left( \bar{R}_q^{(P)} \right)^2 + \frac{2}{3} \cdot \left( \bar{R}_q^{(B)} \right)^2 - \bar{R}_q^2. \quad (8) \end{aligned}$$

We split each of the variance and average terms into texture and motion contributions. We can easily generalize (7) and (8) for arbitrary GoP structures by adjusting the coefficients to represent the fractions of each frame type in the GoP. The general formula for the VD curve model that combines the variabilities of all frame types, or *combined general VD curve model*, is therefore as shown in (9) at the bottom of the page. We can

$$CoV_q = \sqrt{\frac{a^{(I)} \sigma_q^{2(I)} + a^{(P)} \sigma_q^{2(P)} + a^{(B)} \sigma_q^{2(B)} + a^{(I)} \left( \bar{R}_q^{(I)} \right)^2 + a^{(P)} \left( \bar{R}_q^{(P)} \right)^2 + a^{(B)} \left( \bar{R}_q^{(B)} \right)^2}{\left[ a^{(I)} \bar{R}_q^{(I)} + a^{(P)} \bar{R}_q^{(P)} + a^{(B)} \bar{R}_q^{(B)} \right]^2} - 1} \quad (9)$$

rewrite this expression as a function of the coefficients of variation of the I-, P-, and B-frames as shown in (10) at the bottom of the page.

Note that this VD curve model (10) is general in that arbitrary models can be employed for the elementary texture and motion bit statistics (mean, variance, covariance) in (4) and (5) which are in turn inserted in (10). As an instantiation of the texture and motion bit models we develop in the following sections linear and quadratic models for the elementary texture and motion bit statistics. When these linear and quadratic model are employed, we are able to show that the complex concave VD curve shape is the combined result of simpler linear and quadratic curves.

#### IV. TEXTURE AND MOTION BIT STATISTICS MODELS

##### A. Quadratic Models for Texture Bit Statistics and Covariance

In [3], a quadratic rate-distortion model is devised for rate control and this model was adopted as part of -4 VM5.0. The model is formulated with  $a$  and  $b$  as the model parameters:

$$R_q = a \cdot q^{-1} + b \cdot q^{-2}. \quad (11)$$

In this paper, we employ (11) to model the average number of texture bits  $\bar{R}_{q,t}$  in (4), since the average texture bits as a function of  $q$  represent a rate-distortion curve. We also show that (11) accurately models  $\sigma_{q,t}^2$  and is adequate for modeling  $cov_q(t, m)$ . Since the absolute values of the modeled means and variances vary widely, we assess the modeling accuracy with the relative RMSE (RRMSE) values, i.e., the RMSE values divided by the average of the actual data that is modeled on the quantizer scale interval  $q \in [10, 30]$ . The RMSE values are computed based on the difference between the quadratic models and the actual values of the average texture bits, texture variances, and covariances (motion, texture) of the I-, P-, and B-frames.

All model parameters are obtained from the statistics (average, (co)variance) corresponding to the encodings with quantization scales  $q_1 = 10$  and  $q_2 = 30$ . These two  $q$  values are chosen because they include the ‘‘hump’’ of the VD curve and result in a reasonably accurate prediction of the VD curves. Second, the modeling accuracy in the  $q \leq 10$  range is less of a concern to the application of low to medium bit rate video streaming, although we will illustrate in Section IV-D that the VD curve modeling extends towards small  $q$  or high bit rates as well. In Section V, we analyse the impact of the two  $q$  values. The RRMSE values obtained in the described manner across all scenes from Table I and the *Star Wars V* segment are presented in Table II (ignore the motion related table columns denoted by variables with a subscript  $m$  for now). Overall, we

conclude from the table that the quadratic models match the average and variance statistics curves for the texture bits well, while adequately approximating the covariance curves of the P- and B-frames.

Next, we explain a method for estimating the model parameters  $a$  and  $b$ . Let  $X_1$  and  $X_2$  represent  $\bar{R}_{q,t}$ ,  $\sigma_{q,t}^2$ , or  $cov_q(t, m)$  corresponding to two quantization scales  $q_1$  and  $q_2$ . The quadratic model parameters  $a$  and  $b$  from (11) are obtained by solving the following system of equations:

$$X_1 = a \cdot q_1^{-1} + b \cdot q_1^{-2} \quad \text{and} \quad X_2 = a \cdot q_2^{-1} + b \cdot q_2^{-2}. \quad (12)$$

The solution to these equations is:

$$a = \frac{q_1^2 \cdot X_1 - q_2^2 \cdot X_2}{q_1 - q_2}$$

and

$$b = \frac{q_1^2 \cdot q_2 \cdot X_1 - q_2^2 \cdot q_1 \cdot X_2}{q_2 - q_1}. \quad (13)$$

##### B. Linear Models for Motion Bits Statistics

We observe in Fig. 3 that the average number of motion bits follows a linear trend as a function of  $q$ . Hence, we propose a linear model with  $c$  and  $d$  as the model parameters:

$$\bar{R}_{q,m} = c \cdot q + d. \quad (14)$$

The model parameters can be estimated easily by solving a system of two equations as before for the quadratic model:

$$X_1 = c \cdot q_1 + d \quad \text{and} \quad X_2 = c \cdot q_2 + d. \quad (15)$$

The solution to these equations is:

$$c = \frac{X_1 - X_2}{q_1 - q_2} \quad \text{and} \quad d = \frac{q_1 \cdot X_2 - q_2 \cdot X_1}{q_1 - q_2}. \quad (16)$$

Similar to the quadratic models, the linear model parameters are obtained from the motion bits statistics (average, variance) corresponding to the encodings with quantization scales  $q = 10$  and  $q = 30$ . The last constituent of (4) to be modeled is the variance of the motion bits  $\sigma_{q,m}^2$ . With reasoning similar to above, which is detailed in [33], we find that the linear model is also the most appropriate for  $\sigma_{q,m}^2$ .

The RRMSE values for the linear motion bits statistics models for all scenes from Table I and for the *Star Wars V* segment are presented in Table II for the P- and B-frames. We empirically conclude that the linear models fit the average and the variance of the number of motion bits well. Graphical

$$CoV_q = \sqrt{\frac{a^{(I)} \left( \bar{R}_q^{(I)} \right)^2 \left[ \left( CoV_q^{(I)} \right)^2 + 1 \right] + a^{(P)} \left( \bar{R}_q^{(P)} \right)^2 \left[ \left( CoV_q^{(P)} \right)^2 + 1 \right] + a^{(B)} \left( \bar{R}_q^{(B)} \right)^2 \left[ \left( CoV_q^{(B)} \right)^2 + 1 \right]}{\left[ a^{(I)} \bar{R}_q^{(I)} + a^{(P)} \bar{R}_q^{(P)} + a^{(B)} \bar{R}_q^{(B)} \right]^2} - 1} \quad (10)$$

TABLE II  
RRMSE (%) VALUES FOR QUADRATIC MODELS OF TEXTURE BIT STATISTICS AND COVARIANCE AND LINEAR MODELS OF MOTION BIT STATISTICS

Video	Scene	Mot.	$\bar{R}_t^{(I)}$	$\sigma_t^{2(I)}$	$\bar{R}_t^{(P)}$	$\bar{R}_m^{(P)}$	$\sigma_t^{2(P)}$	$\sigma_m^{2(P)}$	$cov^{(P)}$	$\bar{R}_t^{(B)}$	$\bar{R}_m^{(B)}$	$\sigma_t^{2(B)}$	$\sigma_m^{2(B)}$	$cov^{(B)}$
S.W. V	N/A		1.79	6.41	3.32	1.79	4.63	2.54	11.32	4.12	2.84	18.01	6.76	10.82
Footb.	298	I	2.39	12.88	9.82	30.38	2.87	5.42	7.06	22.46	32.64	32.45	7.43	25.69
S.W. IV	274	I	0.73	9.66	12.87	7.12	12.05	18.92	5.96	21.40	5.53	28.46	10.97	39.14
Term.	384	I	1.54	12.69	11.56	13.83	16.08	10.76	6.68	23.58	7.13	27.23	8.28	11.91
Mean		I	1.55	11.74	11.42	17.11	10.34	11.70	6.56	22.48	15.10	29.38	8.89	25.58
Footb.	299	II	2.36	8.84	3.61	1.75	8.72	4.63	14.13	4.87	3.41	20.07	7.60	12.84
S.W. IV	117	II	12.99	5.82	16.00	2.78	12.96	8.18	8.53	23.40	8.95	25.61	16.03	26.81
Term.	462	II	0.52	12.62	5.59	2.36	7.09	4.96	11.72	3.27	3.13	11.06	11.86	11.75
Mean		II	5.29	9.09	8.40	2.30	9.59	5.92	11.46	10.51	5.16	18.91	11.83	17.13
Footb.	557	III	1.11	4.12	2.12	1.01	4.57	9.62	18.46	7.02	4.34	3.98	15.03	8.68
S.W. IV	115	III	2.15	4.31	2.76	2.91	19.95	6.77	8.27	5.10	8.85	27.25	31.44	8.10
Term.	628	III	0.74	8.42	4.63	1.32	7.75	5.68	10.26	2.24	1.75	19.40	4.62	3.95
Mean		III	1.33	5.62	3.17	1.75	10.76	7.36	12.33	4.78	4.98	16.88	17.03	6.91
Footb.	184	IV	0.69	7.81	4.24	0.68	5.52	2.72	8.35	3.05	1.19	3.48	3.04	6.42
S.W. IV	165	IV	3.35	20.34	2.28	1.62	7.69	5.89	13.93	3.60	0.70	6.34	4.26	12.84
Term.	262	IV	2.54	3.92	6.59	1.84	10.74	10.05	8.47	5.28	1.20	17.42	2.96	13.00
Mean		IV	2.19	10.69	4.37	1.38	7.98	6.22	10.25	3.97	1.03	9.08	3.42	10.75
Footb.	336	V	1.41	6.87	6.48	0.35	8.67	9.15	35.13	6.54	0.75	4.72	8.15	29.28
S.W. IV	632	V	4.94	10.50	1.63	1.46	6.74	24.54	15.40	2.08	0.98	3.81	10.76	10.98
Term.	441	V	0.38	3.34	6.86	0.57	3.61	4.74	10.45	5.18	4.16	5.83	4.98	11.99
Mean		V	2.24	6.90	4.99	0.79	6.34	12.81	20.33	4.60	1.96	4.79	7.97	17.41
Mean		I-V	2.52	8.81	6.47	4.66	9.00	8.80	12.19	9.27	5.65	15.81	9.83	15.56

illustrations of the linear and quadratic models, which we can not include here due to space constraints, are provided in [33].

Now that we have developed the models for the texture and motion bits statistics, we employ them in the VD-I, VD-P, and VD-B models from Section III-B and subsequently in the combined VD model from Section III-C.

### C. Evaluation of VD Curve Model for I-, P-, and B-Frames

In Section III-B, we developed VD frame models as given by (4) and (5). The VD-P and VD-B models can be reformulated as a function of the quantization scale  $q$  and ten model parameters, two parameters for each linear and quadratic model:

$$\sigma_{q,t}^2 = a_1/q + b_1/q^2 \quad (17)$$

$$cov_q(t, m) = a_2/q + b_2/q^2 \quad (18)$$

$$\bar{R}_{q,t} = a_3/q + b_3/q^2 \quad (19)$$

$$\sigma_{q,m}^2 = c_1 \cdot q + d_1 \quad (20)$$

$$\bar{R}_{q,m} = c_2 \cdot q + d_2 \quad (21)$$

$$CoV_q^{(P,B)} = \frac{\sqrt{a_1/q + b_1/q^2 + c_1 \cdot q + d_1 + 2 \cdot (a_2/q + b_2/q^2)}}{a_3/q + b_3/q^2 + c_2 \cdot q + d_2} \quad (22)$$

Analogously, the VD-I model includes four model parameters and is given by:

$$\sigma_{q,t}^2 = a_1/q + b_1/q^2 \quad (23)$$

$$\bar{R}_{q,t} = a_2/q + b_2/q^2 \quad (24)$$

$$CoV_{small\ q}^{(P,B)} = CoV_q^{(I)} = \frac{a_1/q + b_1/q^2}{a_2/q + b_2/q^2} \quad (25)$$

From (22) and (25), it is clear that the *complex concave VD curve shape of the individual frame types is the combined result of simpler linear and quadratic curves for the averages, variances, and covariances of the texture and motion bits.*

In Figs. 4 and 5, the actual VD curves for the P- and B-frames from the *Star Wars V* video segment are compared with the VD models estimated from encoding settings  $q_1 = 10$ ,  $q_2 = 30$  and employing (22). The models match the actual curves well for  $10 \leq q \leq 30$  and capture the concave VD curve shape. The VD-P and VD-B models are also an accurate representation for small  $q$  or equivalently the highest qualities. In Fig. 6, the VD-I model and the actual VD curve for the I-frames are depicted. The model matches the VD curve well for  $10 \leq q \leq 30$ . However, the VD-I model becomes inaccurate for  $q < 10$ . Table III enumerates the RRMSE values of the VD-I, VD-P, and VD-B models corresponding to all video scenes in Table I and the *Star Wars V* segment, computed for the quantization scale range  $10 \leq q \leq 30$ . We empirically conclude that the VD curve models for the I-, P-, and B-frames offer a good prediction of the actual VD curves.

In Figs. 7 and 8, the “small  $q$ ” models (5) for the P- and B-frames from the *Star Wars V* segment are depicted (models use  $q_1 = 10$  and  $q_2 = 30$ ). They are adequate approximations for the respective VD curves in the small  $q$  range. From this, it can also be seen that, in the small  $q$  range, the motion statistics and the covariance do not influence the VD curve shape or equivalently there is no influence on the bit rate variability. On the other hand, in the range  $10 \leq q \leq 30$ , the statistics of the motion bits need to be modeled, as well as the covariance, otherwise the curve based on the texture statistics alone deviates strongly from the actual VD curve.

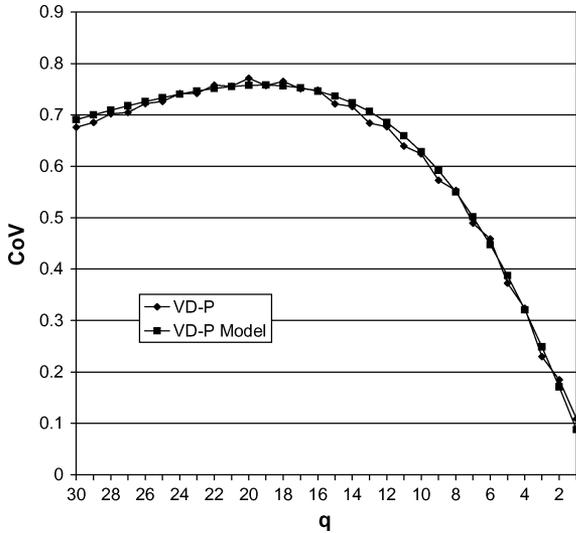


Fig. 4. Comparison of actual VD-P curve with corresponding model.

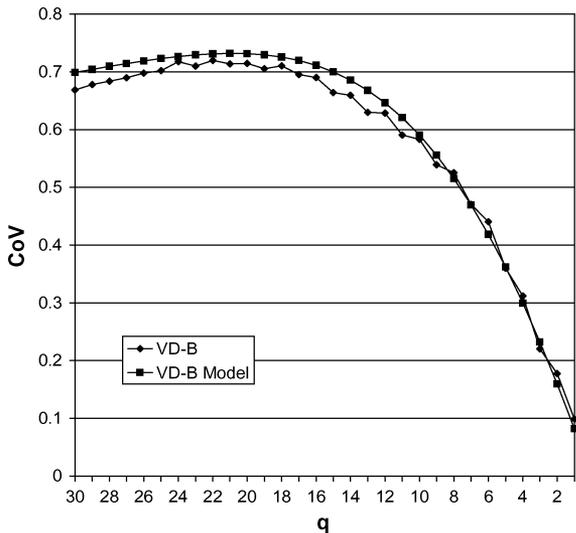


Fig. 5. Comparison of actual VD-B curve with corresponding model.

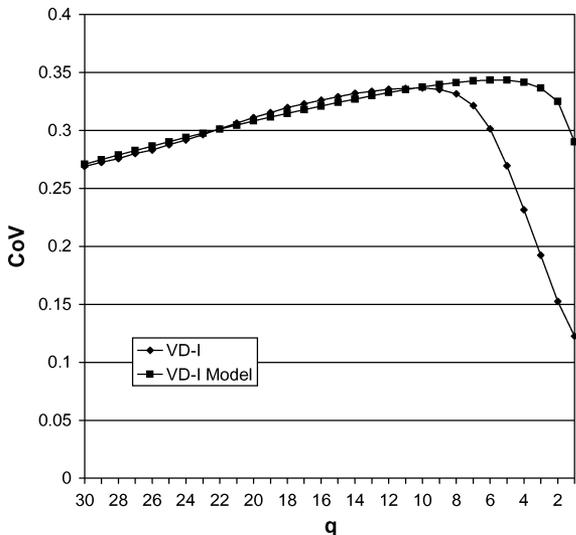


Fig. 6. Comparison of actual VD-I curve with corresponding model.

TABLE III  
RRMSE (%) VALUES FOR VD CURVE MODELS OF I-, P-, AND B-FRAMES

Video	Scene	Mot. Act.	VD-I	VD-P	VD-B
<i>S.W. V</i>	N/A		1.07	1.47	3.45
<i>Footb.</i>	298	I	4.72	14.70	11.17
<i>S.W. IV</i>	274	I	3.63	5.63	4.82
<i>Term.</i>	384	I	8.80	7.04	5.24
<i>Mean</i>		I	4.18	9.12	7.08
<i>Footb.</i>	299	II	1.61	2.07	2.94
<i>S.W. IV</i>	117	II	8.88	3.50	8.58
<i>Term.</i>	462	II	8.08	2.37	6.18
<i>Mean</i>		II	6.19	2.65	5.90
<i>Footb.</i>	557	III	2.79	3.10	4.16
<i>S.W. IV</i>	115	III	3.40	2.85	4.89
<i>Term.</i>	628	III	3.05	4.76	8.59
<i>Mean</i>		III	3.08	3.57	5.88
<i>Footb.</i>	184	IV	3.68	0.76	2.02
<i>S.W. IV</i>	165	IV	7.38	4.23	2.66
<i>Term.</i>	262	IV	2.34	6.27	9.63
<i>Mean</i>		IV	4.47	3.75	4.77
<i>Footb.</i>	336	V	3.65	2.76	5.40
<i>S.W. IV</i>	632	V	7.63	4.57	2.25
<i>Term.</i>	441	V	3.42	4.55	5.68
<i>Mean</i>		V	4.90	3.96	4.44
<i>Mean</i>		I-V	4.56	4.61	5.61

#### D. Evaluation of Combined VD Curve Model

Fig. 9 depicts the combined VD model for the *Star Wars V* video segment and illustrates that the combined VD curve model accurately approximates the actual VD curve of this sequence. In Table IV, we compare the combined VD curve model with the piecewise approximation model applied in [1] for all scenes from Table I using the RRMSE. For the piecewise modeling approach, we present the RRMSE values for approximations obtained from respectively 2, 3, and 4 quantization scales. The RRMSE values are in all cases significantly lower for the combined VD model as compared to the piecewise model employing two quantization scales. For ten out of the fifteen video scenes, the RRMSE values of the combined VD curve model are smaller than the RRMSE values of the piecewise model obtained from three quantization scales, and for about half of the scenes the combined VD curve model outperforms the piecewise RRMSE values obtained from four scales. We conclude from the overall mean that the approximation performance of the combined VD curve model in the quantization scale interval  $10 \leq q \leq 30$ , based on quantization scales  $q_1 = 10$  and  $q_2 = 30$ , is comparable to the piecewise model employing three quantization scales.

To further investigate the accuracy of the combined model approach, we compare the models for the average frame size (7) and standard deviation of the frame size (8) with the corresponding piecewise models in Table V. The table gives the mean RRMSE values across the individual scenes for each of the motion classes considered in Table IV as well as the RRMSE for the *Star Wars V* video sequence (1000 QCIF frames), a new sequence formed by concatenating all fifteen scenes of all motion classes (3395 QCIF frames), and an excerpt from the

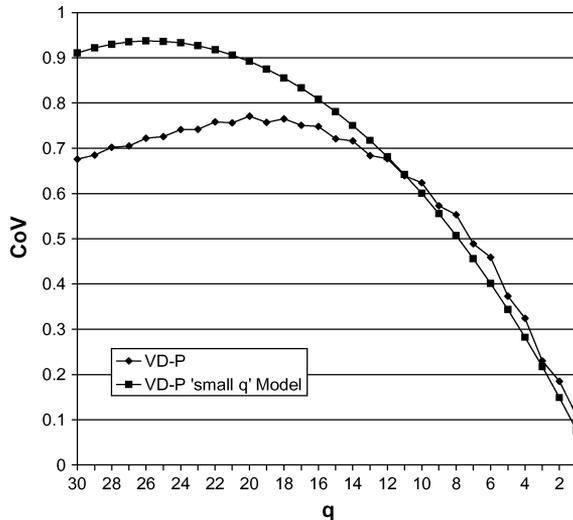
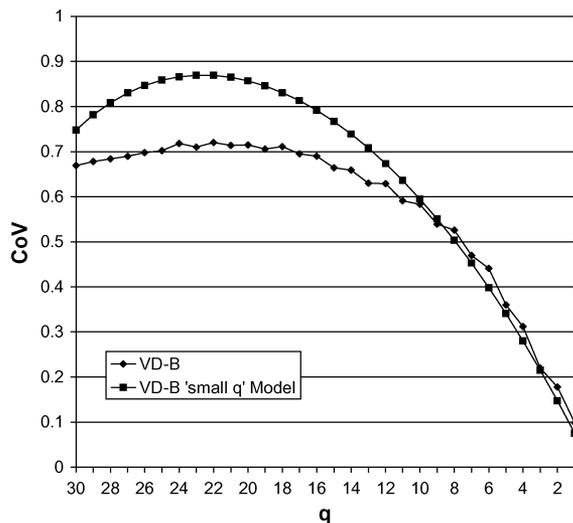
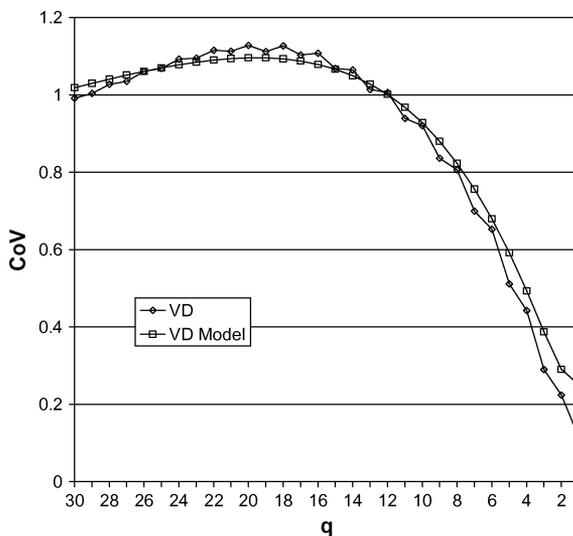
Fig. 7. Actual VD-P curve and the “small  $q$ ” model.Fig. 8. Actual VD-B curve and the “small  $q$ ” model.Fig. 9. VD curve for all frames and the combined VD model for the *Star Wars V* fragment.

TABLE IV  
COMPARISON BETWEEN PIECEWISE APPROXIMATION MODEL (P)  
FROM [1] WITH 2, 3, 4 SAMPLES AND COMBINED VD MODEL  $CoV_q$   
(RRMSE (%) VALUES)

Video	Scene	Mot. Act.	P-2	P-3	P-4	Comb. VD
<i>Footb.</i>	298	I	7.80	3.66	3.72	6.28
<i>S.W. IV</i>	274	I	25.38	7.09	4.93	6.54
<i>Term.</i>	384	I	18.57	4.35	3.55	5.29
<i>Mean</i>		I	17.25	5.03	4.06	6.04
<i>Footb.</i>	299	II	10.48	3.01	2.82	2.52
<i>S.W. IV</i>	117	II	12.64	4.04	4.02	2.58
<i>Term.</i>	462	II	10.91	3.72	2.23	3.28
<i>Mean</i>		II	11.34	3.59	3.02	2.79
<i>Footb.</i>	557	III	16.77	4.80	2.37	1.86
<i>S.W. IV</i>	115	III	12.63	3.07	2.49	1.89
<i>Term.</i>	628	III	13.14	4.59	2.58	2.47
<i>Mean</i>		III	14.18	4.15	2.48	2.07
<i>Footb.</i>	184	IV	8.26	2.59	1.72	1.53
<i>S.W. IV</i>	165	IV	14.44	4.82	2.41	1.64
<i>Term.</i>	262	IV	9.10	3.17	1.93	5.80
<i>Mean</i>		IV	10.60	3.53	2.02	2.99
<i>Footb.</i>	336	V	7.15	2.72	2.34	3.07
<i>S.W. IV</i>	632	V	12.70	6.23	1.86	3.28
<i>Term.</i>	441	V	8.83	2.03	1.48	4.78
<i>Mean</i>		V	9.56	3.66	1.89	3.71
<i>Mean</i>		I-V	12.59	3.99	2.70	3.52

*NBC 12 News* (3600 CIF frames). We observe that for the individual scenes, the combined model generally compares favorably with the piecewise approximation approach with three samples. For the long sequences, which include a variety of scenes and motion activities and are typical for video streaming applications, the combined model compares favorably with the piecewise model with four samples. This indicates that the combined model tends to perform particularly well for video with diverse motion activity content. For the diverse content, the larger number of model parameters that separately capture the elementary motion and texture frame size statistics and are then combined to model the overall frame size statistics tend to give more accurate characterizations than the piecewise model which only considers the total frame sizes and is oblivious to the underlying video content.

To illustrate the impact of our combined model in an application scenario, we consider stream admission control for bufferless statistical multiplexing on a link with capacity  $C$ . We determine the maximum number of video streams  $J_{\max}$  that can be supported subject to a limit  $\varepsilon$  on the loss probability. We consider the Central Limit based admission control outlined in [1], which models the aggregate traffic load as a Normal random variable with mean  $\sum_{j=1}^J \bar{R}_q(j)$  and variance  $\sum_{j=1}^J \sigma_q^2(j)$ . The mean  $\bar{R}_q(j)$  and the variance  $\sigma_q^2(j)$  as a function of  $q$  can be obtained using the corresponding combined or piecewise models for sequence  $j$ . We compare the accuracies of the  $J_{\max}$  admission decisions based on the combined and piecewise models in Table VI for the *Star Wars V* and concatenated scenes sequences for a link with  $C = 5$  Mbps, and for the *NBC 12 News* sequence for a link with  $C = 20$  Mbps. We find that the RRMSE values of the combined modeling are significantly lower than

TABLE V  
COMPARISON BETWEEN PIECEWISE APPROXIMATION (P) AND COMBINED MODEL FOR FRAME SIZE MEAN ( $\bar{R}_q$ ) AND STANDARD DEVIATION ( $\sigma_q$ ) (RRMSE (%) VALUES)

Means	P-2 $\bar{R}_q$	P-3 $\bar{R}_q$	P-4 $\bar{R}_q$	Comb. $\bar{R}_q$	P-2 $\sigma_q$	P-3 $\sigma_q$	P-4 $\sigma_q$	Comb. $\sigma_q$
I	16.37	6.28	4.55	7.25	4.06	2.06	1.13	1.21
II	12.60	4.66	2.83	4.73	7.74	3.74	2.10	4.18
III	13.37	5.28	2.95	2.53	3.27	1.42	0.72	1.04
IV	9.77	4.47	2.58	2.42	4.29	2.11	0.82	1.35
V	7.04	3.29	1.87	2.82	3.72	1.63	1.05	1.70
I-V	11.83	4.79	2.96	3.95	4.62	2.19	1.17	1.90
S.W. V	15.68	7.57	4.07	1.49	5.85	2.76	1.35	1.64
Concat.	14.11	5.86	3.21	1.60	2.24	1.00	0.32	0.67
News	11.08	4.60	2.11	0.71	3.09	1.40	0.50	0.18
Mean	13.62	6.01	3.13	1.27	3.73	1.72	0.72	0.83

TABLE VI  
COMPARISON OF  $J_{\max}$  ESTIMATES BASED ON PIECEWISE APPROXIMATION (P) WITH 2, 3, AND 4 SAMPLES AND COMBINED MODEL (RRMSE (%) VALUES) FOR PERMITTED LOSS PROBABILITIES OF  $\epsilon = 10^{-7}$  AND  $10^{-3}$

Video	P-2 ( $10^{-7}$ )	P-3 ( $10^{-7}$ )	P-4 ( $10^{-7}$ )	Comb. ( $10^{-7}$ )	P-2 ( $10^{-3}$ )	P-3 ( $10^{-3}$ )	P-4 ( $10^{-3}$ )	Comb. ( $10^{-3}$ )
S.W. V	5.41	3.02	2.70	1.58	7.71	3.07	2.38	1.36
Concat.	6.95	3.71	2.94	1.28	8.79	2.63	2.63	2.19
News	5.47	2.26	2.26	0.68	6.92	2.37	1.89	1.09
Mean	5.94	3.00	2.63	1.18	7.80	2.69	2.30	1.54

the values of the piecewise approach using four samples, indicating that the combined model leads to significantly more accurate admission decisions. This result is consistent with the significantly more accurate modeling of  $\bar{R}_q$  by the combined model compared to the piecewise model with four samples and the comparable accuracy for  $\sigma_q$  of the two models for the long sequences in Table V.

## V. SENSITIVITY OF VD CURVE MODEL

In this section, we study the impact of the two quantization scales,  $q_1$  and  $q_2$ , that are used for encoding the video sequence and subsequently for estimating the averages, variances, and covariances of the number of texture and motion bits. The linear and quadratic models that we have developed, could strongly depend on the choice of  $q_1$  and  $q_2$ . For example, if  $q_1$  and  $q_2$  are not spread out over a significant portion of the quantization scale range  $1 \leq q \leq 30$ , then it is likely that a modeling error will be substantial for quantization scales that are located far from  $q_1$  and  $q_2$ . Furthermore, the ‘‘hump’’ shape is influenced by both the motion and the texture statistics, as observed earlier. Therefore, the motion statistics need to be modeled accurately by choosing at least one  $q$  value in the higher compression ratio region (large  $q$ ). Otherwise, if  $q_1$  and  $q_2$  would be selected in the low compression ratio region, then we would estimate motion statistics in a region where the texture information dominates the motion information, resulting in a good approximation of the motion bits statistics in a region where motion information is insignificant for determining the VD curve shape. From these considerations, we can already formulate two high-level recommendations for the choice of  $q_1$  and  $q_2$ : (i) there should be a reasonable interval size  $\Delta q = q_1 - q_2$ , and (ii) at least one  $q$  should be chosen in the high compression ratio region where the ‘‘hump’’ is typically situated.

TABLE VII  
QUANTIZATION SCALE INTERVALS  $\Delta q$  AND POSSIBLE CHOICES FOR  $q_1$  AND  $q_2$

$\Delta q$	$q_1$	$q_2$
20	1	20
	5	25
	10	30
30	1	30

An important unknown is the possibility that the choice of  $q_1$  and  $q_2$  depends on the motion activity in the video scene, since the ‘‘hump’’ amplitude is influenced by the motion activity [1]. It seems reasonable to examine whether there could be better choices for  $q_1$  and  $q_2$  than 10 and 30 (which we used in the preceding section), depending on the motion activity, and resulting in more accurate concave VD curve models. In our sensitivity analysis of the choices for  $q_1$  and  $q_2$ , we examine interval sizes  $\Delta q \in \{20, 30\}$ , and choices for  $q_1$  and  $q_2$  from the range  $1 \leq q \leq 30$ , as enumerated in Table VII. These four possible combinations cover the entire quantization scale interval. The encodings are followed by estimating the texture and motion statistics, and subsequently by calculating the VD curve models for the P-, B-, and I-frames. We have omitted the interval size  $\Delta q = 10$ , since our experiments indicate that the VD curve models for  $\Delta q = 10$  are poor approximations of the actual VD curves for most video scenes.

We have computed the RMSE for each  $q_1$  and  $q_2$  choice from Table VII, for each video scene from Table I. The RMSE is computed for the entire range of the quantization scale. Table VIII enumerates the average RMSE values from each of the five motion classes. We observe sensitivities to the choice of  $q_1$  and  $q_2$  for the low motion activity classes I and II for which the VD curve hump, which is generally between the middle and upper end of the  $q$  range, is most pronounced. A  $q_2$  that is removed

TABLE VIII  
SENSITIVITY ANALYSIS OF QUANTIZATION SCALES USED FOR MODEL PARAMETER ESTIMATION (MEAN RMSE VALUES)

Mot. Act.	$q_1 = 1, q_2 = 30$	$q_1 = 1, q_2 = 20$	$q_1 = 10, q_2 = 30$	$q_1 = 5, q_2 = 25$
I	0.0733	0.1299	0.0479	0.2318
II	0.1543	0.1750	0.0740	0.0705
III	0.0877	0.0746	0.0925	0.0709
IV	0.0449	0.0367	0.0321	0.0307
V	0.0347	0.0266	0.0399	0.0400
I-V	0.0790	0.0886	0.0573	0.0888

from the upper end of the  $q$  range, e.g.,  $q_2 = 20$  or  $25$ , tends to capture the VD behavior somewhat less accurately for low motion scenes. For the higher motion activity classes III, IV, and V, we observe relatively little differences between the  $q_1$  and  $q_2$  choices. Overall, we observe that the quantization scale choice  $q_1 = 10$  and  $q_2 = 30$  has the lowest average RMSE across all motion activity classes, and has an average RMSE lower than 0.1 for each motion class, whereas the other choices have RMSE values exceeding 0.1 for low motion scenes. From this sensitivity analysis we can recommend that a good choice for the two quantization scales is  $q_1 = 10$  and  $q_2 = 30$ , and note that it is generally possible to adapt the  $q_1$  and  $q_2$  choice to the scene motion content if scene segmentation is employed.

## VI. CONCLUSION

We have modeled the bit rate variability-distortion (VD) curve for video streaming. The VD curve has a typical concave or “hump” shape for -4 VBR encodings as previously observed in [1]. We have refined the coefficient of variation  $CoV$  into the  $CoV$  of texture and motion information, since we have found that the motion information is an important constituent at medium to high compression ratios. Furthermore, the texture and motion information statistics are different functions of the quantization scale  $q$  and therefore splitting the frame sizes into texture and motion bits is warranted. Based on elementary texture and motion bit statistics, we have formulated VD curve models for the different frame types (I, P, and B) as well as a general VD curve model for a video sequence.

We modeled the average and variance of the texture bits using a quadratic model based on the rate-distortion model employed for rate control in -4 [3]. We also modeled the covariance of the texture and motion bits by this quadratic model. We modeled the average and the variance of the motion bits by a linear model. We have thus extended [3] where a quadratic model was used for the rate-distortion function of the entire frame size. Employing these linear and quadratic models in our general VD curve model results overall in good predictions of the actual VD curves and significantly improves upon the piecewise modeling in [1]. The texture and motion based VD curve model for two quantization scales has a modeling accuracy that is comparable to the piecewise model for three quantization scales.

A future research direction is adapting the VD curve modeling approach developed in this paper to the new H.264/AVC [34] video encoder. A possible approach is based on the quadratic rate-distortion model proposed in [35], which employs the relationship between the new quantization parameter  $QP$  and the quantization step in the H.264/AVC encoder.

## APPENDIX A

### EVALUATION OF VARIANCE $\sigma_q^2$

With the notation of Section III-C we obtain for the overall variance  $\sigma_q^2$  for a particular quantization scale  $q$ :

$$\sigma_q^2 = \frac{1}{12N} \left[ \sum_{i=1}^N \left( X_{q,i}^{(I)} - \bar{R}_q \right)^2 + \sum_{i=1}^{3N} \left( X_{q,i}^{(P)} - \bar{R}_q \right)^2 + \sum_{i=1}^{8N} \left( X_{q,i}^{(B)} - \bar{R}_q \right)^2 \right] \quad (26)$$

$$= \frac{1}{12N} \sum_{i=1}^N \left( \left\{ X_{q,i}^{(I)} - \bar{R}_q^{(I)} \right\} + \left\{ \bar{R}_q^{(I)} - \bar{R}_q \right\} \right)^2 + \frac{1}{12N} \sum_{i=1}^{3N} \left( \left\{ X_{q,i}^{(P)} - \bar{R}_q^{(P)} \right\} + \left\{ \bar{R}_q^{(P)} - \bar{R}_q \right\} \right)^2 + \frac{1}{12N} \sum_{i=1}^{8N} \left( \left\{ X_{q,i}^{(B)} - \bar{R}_q^{(B)} \right\} + \left\{ \bar{R}_q^{(B)} - \bar{R}_q \right\} \right)^2 \quad (27)$$

$$= \frac{1}{12N} \left[ \sum_{i=1}^N \left( X_{q,i}^{(I)} - \bar{R}_q^{(I)} \right)^2 + \sum_{i=1}^{3N} \left( X_{q,i}^{(P)} - \bar{R}_q^{(P)} \right)^2 + \sum_{i=1}^{8N} \left( X_{q,i}^{(B)} - \bar{R}_q^{(B)} \right)^2 \right] + \frac{1}{6N} \left[ \left( \bar{R}_q^{(I)} - \bar{R}_q \right) \sum_{i=1}^N \left( X_{q,i}^{(I)} - \bar{R}_q^{(I)} \right) \right] + \frac{1}{6N} \left[ \left( \bar{R}_q^{(P)} - \bar{R}_q \right) \sum_{i=1}^{3N} \left( X_{q,i}^{(P)} - \bar{R}_q^{(P)} \right) \right] + \frac{1}{6N} \left[ \left( \bar{R}_q^{(B)} - \bar{R}_q \right) \sum_{i=1}^{8N} \left( X_{q,i}^{(B)} - \bar{R}_q^{(B)} \right) \right] + \frac{1}{12} \left[ \left( \bar{R}_q^{(I)} - \bar{R}_q \right)^2 + 3 \cdot \left( \bar{R}_q^{(P)} - \bar{R}_q \right)^2 + 8 \cdot \left( \bar{R}_q^{(B)} - \bar{R}_q \right)^2 \right]. \quad (28)$$

The  $1/12N$  terms can clearly be rewritten as the variances of the three frame types (I-P-B). The  $1/6N$  terms each sum over the individual frame sizes about the mean for each frame type, which is the first central moment. Hence, all  $1/6N$  terms equal zero. The last terms with coefficient  $1/12$  need some expansion to simplify them. This results in (8) for the overall variance of the frame sizes.

## ACKNOWLEDGMENT

The authors are grateful to the thoughtful comments of the anonymous reviewers, which helped to significantly improve this manuscript.

## REFERENCES

- [1] P. Seeling and M. Reisslein, "The rate variability-distortion (VD) curve of encoded video and its impact on statistical multiplexing," *IEEE Trans. Broadcasting*, vol. 51, no. 4, pp. 473–492, Dec. 2005.
- [2] ISO/IEC JTC 1/SC 29/WG 11 N2802, in *Information technology—generic coding of audio-visual objects—part 2: Visual, final proposed draft amendment 1*, Geneva, July 1999.
- [3] T. Chiang and Y.-Q. Zhang, "A new rate control scheme using quadratic rate distortion model," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 7, no. 1, Feb. 1997.
- [4] A. Ortega and K. Ramachandran, "Rate-distortion methods for image and video compression," *IEEE Signal Processing Magazine*, vol. 15, no. 6, pp. 23–50, Nov. 1998.
- [5] H.-M. Hang and J.-J. Chen, "Source model for transform video coder and its application—Part I: Fundamental theory," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 7, no. 2, pp. 287–298, Apr. 1997.
- [6] C. Cai, R. Ding, and S. K. Mitra, "New bit-rate control models for image and video transmission," in *Proc. of the IEEE 2002 International Conference on Communications, Circuits and Systems and West Sino Expositions*, Jun. 2002, vol. 2, pp. 958–962.
- [7] J. Sun, W. Gao, D. Zhao, and Q. Huang, "Statistical model, analysis and approximation of rate-distortion function in -4 FGS videos," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 16, no. 4, pp. 535–539, Apr. 2006.
- [8] Z. Zhang, G. Liu, H. Li, and Y. Li, "A novel PDE-based rate-distortion model for rate control," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 15, no. 11, pp. 1354–1364, Nov. 2005.
- [9] W. Ding and B. Liu, "Rate control of video coding and recording by rate-quantization modeling," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 6, no. 1, pp. 12–20, Feb. 1996.
- [10] M. Dai and D. Loguinov, "Analysis of rate-distortion functions and congestion control in scalable internet video streaming," in *Proc. of the Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV)*, Monterey, CA, Jun. 2003, pp. 60–69.
- [11] X. Lu, R. O. Morando, and M. ElZarki, "Understanding video quality and its use in feedback control," in *Proceedings of International Packet Video Workshop 2002*, Pittsburgh, PA, 2002.
- [12] A. Alheraish, S. Alshebeli, and T. Alamri, "A GACS modeling approach for broadcast video," *IEEE Trans. Broadcasting*, vol. 50, no. 2, pp. 132–141, Jun. 2004.
- [13] N. Ansari, H. Liu, Y. Q. Shi, and H. Zhao, "On modeling video traffics," *IEEE Trans. Broadcasting*, vol. 48, no. 4, pp. 337–347, Dec. 2002.
- [14] M. Dai and D. Loguinov, "Analysis and modeling of -4 and H.264 multi-layer video traffic," in *Proceedings of IEEE Infocom*, Mar. 2005, pp. 2257–2267.
- [15] X.-D. Huang, Y.-H. Zhou, and R.-F. Zhang, "A multiscale model for -4 varied bit rate video traffic," *IEEE Trans. Broadcasting*, vol. 50, no. 3, pp. 323–334, Sept. 2004.
- [16] M. M. Krunz and A. M. Makowski, "Modeling video traffic using  $M/G/\infty$  input processes: A compromise between Markovian and LRD models," *IEEE Journal on Selected Areas in Communications*, vol. 16, pp. 733–748, Jun. 1998.
- [17] C. H. Liew, C. K. Kodikara, and A. M. Kondoz, "—encoded variable bit-rate video traffic modelling," *IEE Proceedings Communications*, vol. 152, no. 5, pp. 749–756, Oct. 2005.
- [18] U. K. Sarkar, S. Ramakrishnan, and D. Sarkar, "Modeling full-length video using Markov-modulated gamma-based framework," *IEEE/ACM Trans. Networking*, vol. 11, no. 4, pp. 638–649, Aug. 2003.
- [19] P. Frossard and O. Verscheure, "Joint source/FEC rate selection for quality-optimal -2 video delivery," *IEEE Trans. Image Processing*, vol. 10, no. 12, pp. 1815–1825, Dec. 2001.
- [20] W. Luo and M. ElZarki, "Quality control for VBR video over ATM networks," *IEEE Journal on Selected Areas in Communications*, vol. 15, no. 6, pp. 1029–1039, Aug. 1997.
- [21] Y. Yang and S. S. Hernami, "Rate control for VBR video over ATM: simplification and implementation," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 11, no. 9, pp. 1045–1058, Sept. 2001.
- [22] S. Bakiras and V. O. K. Li, "Maximizing the number of users in an interactive video-on-demand system," *IEEE Trans. Broadcasting*, vol. 48, no. 4, pp. 281–292, Dec. 2002.
- [23] I. Dalgic and F. A. Tobagi, "Performance evaluation of ATM networks carrying constant and variable bit-rate video traffic," *IEEE Journal on Selected Areas in Communications*, vol. 15, no. 6, pp. 1115–1131, Aug. 1997.
- [24] M. Krunz, W. Zhao, and I. Matta, "Scheduling and bandwidth allocation for distribution of archived video in VoD systems," *Telecommunication Systems*, vol. 9, no. 3/4, pp. 335–355, Sept. 1998.
- [25] M. Balakrishnan, R. Cohen, E. Fert, and G. Keesman, "Benefits of statistical multiplexing in multi-program broadcasting," in *Proceedings of Int. Broadcasting Convention*, Sept. 1997, pp. 560–565.
- [26] P. Cuenca, A. Garrido, F. Quiles, and L. Orozco-Barbosa, "An efficient protocol architecture for error-resilient -2 video communications over ATM networks," *IEEE Trans. Broadcasting*, vol. 45, no. 1, pp. 129–140, Mar. 1999.
- [27] J. Jordan and A. Bock, "Analysis, modeling and performance prediction of digital video statistical multiplexing," in *Proceedings of Int. Broadcasting Convention*, Sept. 1997, pp. 553–559.
- [28] G.-M. Muntean, P. Perry, and L. Murphy, "A new adaptive multimedia streaming system for all-IP multi-service networks," *IEEE Trans. Broadcasting*, vol. 50, no. 1, pp. 1–10, Mar. 2004.
- [29] *American National Standard for Telecommunications—Digital Transport of One-way Video Signals—Parameters for Objective Performance Assessment*, ANSI T1.801.03-2003, American National Standards Institute.
- [30] "Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment, Phase II," Aug. 2003 [Online]. Available: <http://www.vqeg.org>
- [31] "Automatic Movie Content Analysis (MoCA), Shot Boundary Detection tool," [Online]. Available: <http://www.informatik.uni-mannheim.de/pi4/projects/MoCA>
- [32] S. Jeannin and A. Divakaran, "—7 visual motion descriptors," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 11, no. 6, pp. 720–724, Jun. 2001.
- [33] G. Van der Auwera, M. Reisslein, and L. Karam, Video texture and motion based modeling of rate variability-distortion (VD) curves (extended version) Tempe, AZ, June 2006 [Online]. Available: <http://trace.eas.asu.edu>, ASU, Tech. Rep.
- [34] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, Jul. 2003.
- [35] Z. Li, W. Gao, F. Pan, S. Ma, K. Lim, G. Feng, X. Lin, S. Rahardja, H. Lu, and Y. Lu, "Adaptive rate control for H.264," *Elsevier Journal of Visual Communication & Image Representation*, vol. 17, no. 2, pp. 376–406, Apr. 2006.



**Geert Van der Auwera** received the electrotechnical civil engineering degree from Vrije Universiteit Brussel (VUB), Brussels, Belgium in 1997. Presently, he is working towards the Ph.D. degree in the Electrical Engineering Dept. of Arizona State University, USA. His research encompasses the analysis and modeling of video traffic characteristics. Until end of 2004, he was scientific advisor for IWT-Flanders, the Institute for the Promotion of Innovation by Science and Technology in Flanders, Belgium. In 2000, he joined IWT-Flanders after performing research at the Electronics and Information Processing Department (ETRO), VUB. His fields of interest are video coding, and video streaming technologies. In 1998, Mr. Van der Auwera's thesis on motion estimation in the wavelet domain received the Barco and IBM prizes by the Fund for Scientific Research—Flanders.



**Martin Reisslein** is an Associate Professor in the Department of Electrical Engineering at Arizona State University (ASU), Tempe. He received the Dipl.-Ing. (FH) degree from the Fachhochschule Dieburg, Germany, in 1994, and the M.S.E. degree from the University of Pennsylvania, Philadelphia, in 1996. Both in electrical engineering. He received his Ph.D. in systems engineering from the University of Pennsylvania in 1998. During the academic year 1994–1995 he visited the University of Pennsylvania as a Fulbright scholar. From July 1998 through October 2000

he was a scientist with the German National Research Center for Information Technology (GMD FOKUS), Berlin and lecturer at the Technical University Berlin. From October 2000 through August 2005 he was an Assistant Professor at ASU.

From 2003 through 2006 he was editor-in-chief of the *IEEE Communications Surveys and Tutorials*. He has served on the Technical Program Committees of *IEEE Infocom*, *IEEE Globecom*, and the *IEEE International Symposium on Computer and Communications*. He has organized sessions at the *IEEE Computer Communications Workshop (CCW)*. He maintains an extensive library of video traces for network performance evaluation, including frame size traces of  $-4$  and H.263 encoded video, at <http://trace.eas.asu.edu>. He is co-recipient of Best Paper Awards of the *SPIE Photonics East 2000—Terabit Optical Networking conference* and the 2006 *IEEE Consumer Communications and Networking Conference*. His research interests are in the areas of Internet Quality of Service, video traffic characterization, wireless networking, optical networking, and engineering education.



**Lina J. Karam** received the B.E. degree in Computer and Communications Engineering from the American University of Beirut in 1989, and the M.S. and Ph.D. degrees in Electrical Engineering from the Georgia Institute of Technology in 1992 and 1995, respectively. She is currently an associate professor in the Electrical Engineering department at Arizona State University. Her research interests are in the areas of image and video processing, image and video coding, error-resilient source coding, digital filtering, and medical imaging.

From 1991 to 1995, Prof. Karam was a graduate research assistant in the Graphics, Visualization, and Usability Center and then in the Department of Electrical Engineering at Georgia Tech. She worked at Schlumberger Well Services on problems related to data modeling and visualization, and in the Signal Processing Department of AT&T Bell Labs on problems in video coding during 1992 and 1994, respectively. Prof. Karam is the recipient of an NSF CAREER Award. She served as the Chair of the IEEE Communications and Signal Processing Chapters in Phoenix in 1997 and 1998. She also served as an associate editor of the *IEEE TRANSACTIONS ON IMAGE PROCESSING* from 1999 to 2003. She is serving as an associate editor of the *IEEE SIGNAL PROCESSING LETTERS*. She is an elected member of the IEEE Circuits and Systems Society's DSP Technical Committee, the IEEE Signal Processing Society's IMDSP Technical Committee, and a member of the IEEE Signal Processing Society's Conference Board. She is also a member of the Signal Processing, and Circuits and Systems societies of the IEEE.