VIDEO TEXTURE AND MOTION BASED MODELING OF RATE VARIABILITY-
DISTORTION (VD) CURVES OF I, P, AND B FRAMES

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

Arizona State University
Dept. of Electrical Engineering
Tempe, AZ 85287-5706
{geert.vanderauwera, reisslein, karam}@asu.edu
http://trace.eas.asu.edu

ABSTRACT
We examine the bit rate variability-distortion (VD) curve of I,
P, and B frames of MPEG-4 VBR encoded video sequences.
We show that the concave VD curve shape at high compres-
sion ratios or large quantization scales, is influenced by both
the texture and the motion information. We use linear and
quadratic models for the texture and motion bits statistics and
device accurate VD curve models. The model parameters are
obtained from statistics that are estimated from two encod-
ings. This work extends our previous work on modeling the
VD curve, which has applications for optimal statistical mul-
tiplexing of VBR streaming video.

1. INTRODUCTION
A recent study [1] has documented the concave shape of the
rate variability-distortion (VD) curve of open-loop (VBR) en-
coded video and proposed a crude piecewise model for the
VD curve. The VD curve is the coefficient of variation
(CoV) (standard deviation normalized by the mean) of the frame size
(in bits) as a function of the quantization scale. In this study,
we build on [1] by examining the underlying effects leading
to the concave VD curve and by developing and validating a
refined VD curve model. More specifically, we develop and
validate quadratic models for the mean and (co)variance of
the texture information (bits) in a frame. We also develop
linear models for the mean and variance of the motion infor-
mation (bits) in a frame, thus extending [2] where a quadratic
model was used for the rate-distortion function of the entire
frame size. We combine the models of the texture and mo-
tion bits statistics in an overall model of the VD curve. We
demonstrate that both the texture and motion bits make sig-
nificant contributions to the overall concave shape of the VD
curve. We find that given the frame size contributions for en-
codings 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 piecewise model in [1] where encodings for four or
more different quantization scales were required, and which
did not provide insights into how the underlying video content
features give rise to the VD curve characteristics.

The VD curve has important implications for statistical
multiplexing as the highest statistical multiplexing gain is typ-
ically achieved at the peak of the VD curve [1]. Thus, the
refined VD curve model proposed in this paper, not only pro-
vides fundamental insights into how the content features tex-
ture and motion in an encoded video contribute to its traffic
variability, but also provide a practical method for estimating
the rate variability from only two sample encodings.

2. VD CURVE MODEL
2.1. Texture vs. Motion Bits
Encoded video frames have as main constituents texture, mo-
tion, and syntax bits. For large quantization scales \( q \), the num-
ber of texture bits is comparable to the motion bits and there-
fore 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 infor-
mation and also by the motion information. We consider the
syntax bits negligible compared to the texture and motion bits,
since they are more than an order of magnitude smaller. This
is illustrated in Fig. 1 for a Star Wars V: The Empire Strikes
Back segment (1000 QCIF frames are used). For the encoding,
we used the MPEG-4 codec (ISO/IEC JTC 1/SC 29/WG
11 N2802, July 1999).

2.2. Texture and Motion VD Curve Model
For a given quantization scale \( q \), let \( \bar{R}_{q,t} \) and \( \sigma_{q,t}^2 \) respectively
denote the average and the variance of the number of texture
bits, and \( \bar{R}_{q,m} \) and \( \sigma_{q,m}^2 \) respectively denote the average
and variance of the number of motion bits. \( \text{cov}_q(t,m) \) represents
the covariance of the texture and motion information. We ob-
serve 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: $R_q = R_{q,t} + R_{q,m}$. We can approximate the $CoV_q$ for the P and B frames by:

$$CoV_q^{(P,B)} = \frac{\sigma_q}{R_q} = \sqrt{\frac{\sigma_{q,t}^2 + \sigma_{q,m}^2 + 2 \cdot cov_q(t,m)}{R_{q,t} + R_{q,m}}}.$$  \hspace{1cm} (1)

For small $q$ values, the motion bits are negligible compared to the texture bits and therefore Eqn. (1) reduces to:

$$CoV_{q_{small}}^{(P,B)} = CoV_q^{(I)} = \frac{\sigma_{q,t}}{R_{q,t}}.$$ \hspace{1cm} (2)

Eqn. (2) is applicable to the I frames as well, since no motion information is present. In the following sections, we formulate models for each of the constituents in Eqns. (1) and (2).

### 2.3. Quadratic Models

In [2], a quadratic rate-distortion model is devised and the rate control algorithm based on this model was adopted as part of MPEG-4 VM5.0. The model is formulated in the following equation, with $a$ and $b$ the model parameters:

$$R_q = a \cdot q^{-1} + b \cdot q^{-2}.$$ \hspace{1cm} (3)

In this paper, we will employ Eqn. (3) to model the average number of texture bits $R_{q,t}$ in Eqn. (1), since the average texture bits as a function of $q$ represent a rate-distortion curve. We also show that Eqn. (3) accurately models $\sigma_{q,t}^2$ and is adequate for modeling $cov_q(t,m)$.

Figs. 2–4 illustrate the quadratic modeling of the Star Wars segment statistics. We empirically conclude that the quadratic models match the average and variance statistics curves well, while adequately approximating the covariance curves. The modeling error of the covariance curves in the $q \leq 10$ range is acceptable, since in this range the covariance values are an order of magnitude smaller than the variance values. All model
parameters are obtained from the statistics corresponding to the encodings with quantization scales $q = 10$ and $q = 30$. In [3], we analyze the sensitivity of the choice of these two $q$ values and illustrate the modeling for many video sequences. A method for estimating the model parameters $a$ and $b$ is explained next.

Let $X_1$ and $X_2$ represent $R_{q,t}$, $\sigma_{q,t}^2$, or $\text{cov}_{q}(t,m)$ corresponding to two quantization scales $q_1$ and $q_2$. The quadratic model parameters $a$ and $b$ from Eqn. (3) are obtained by solving the following system of equations:

$$X_1 = a \cdot q_1^{-1} + b \cdot q_1^{-2}$$
$$X_2 = a \cdot q_2^{-1} + b \cdot q_2^{-2}.$$  

The solution to these equations is:

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

2.4. Linear Models

We observe in Fig. 1 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:

$$R_{q,m} = c \cdot q + d.$$  

The model parameters can be estimated easily by solving a system of two linear equations. Fig. 5 illustrates the linear model for the P, B motion averages of the Star Wars V segment. We observe that the linear model accurately fits the average curves. The last constituent of Eqn. (1) to be modeled is the variance of the motion bits, $\sigma_{q,m}^2$. Fig. 6 depicts $\sigma_{q,m}^2$ for the P and B frames. The linear model is also the most appropriate in this case. Now that we have developed the individual models, we are ready to assemble the VD curve model.

2.5. VD Curve Model

The complete VD curve model (Eqn. (1)) for the P and B frames can be reformulated as a function of $q$ and ten model parameters:

$$\sigma_{q,t}^2 = \frac{a_1}{q} + \frac{b_1}{q^2}$$
$$\text{cov}_{q}(t,m) = \frac{a_2}{q} + \frac{b_2}{q^2}$$
$$R_{q,t} = \frac{a_3}{q} + \frac{b_3}{q^2}$$
$$\sigma_{q,m}^2 = c_1 \cdot q + d_1$$
$$R_{q,m} = c_2 \cdot q + d_2$$

$$\text{CoV}_{q}^{P,B} = \sqrt{\frac{a_3}{q} + \frac{b_3}{q^2} + c_1 \cdot q + d_1 + 2 \left( \frac{a_2}{q} + \frac{b_2}{q^2} \right)}.$$  

Analogously, the VD curve model for small $q$ values and for the I frames is given by:

$$\sigma_{q,t}^2 = \frac{a_1}{q} + \frac{b_1}{q^2}$$
$$R_{q,t} = \frac{a_2}{q} + \frac{b_2}{q^2}$$

$$\text{CoV}_{q}^{I} = \sqrt{\frac{a_1}{q} + \frac{b_1}{q^2}}.$$  

All model parameters are estimated from the two sample encodings as explained in sections 2.3 and 2.4. In Fig. 7, the VD curves for the I, P and B frames are depicted alongside the VD models estimated from encoding settings $q = 10$, $q = 30$. The models match the original 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. The VD-I model matches the VD curve well for $10 \leq q \leq 30$. In Fig. 8, we present the VD curves and models for a scene from the Football sequence. In [3] we present the validation of the VD modeling for many video sequences.
In Figs. 9 and 10, the small $q$ models for the P and B frames from the Star Wars V segment are depicted (models use $q = 10$ and $q = 30$). They are adequate approximations for the respective VD curves in the small $q$ range ($q < 10$). 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.

3. CONCLUSION
We have modeled the bit rate variability-distortion (VD) curves of the I, P and B frames. The coefficient of variation (CoV) of the MPEG-4 encoded frame sizes for P and B frames has been refined into the CoV of texture and motion information. We found that the VD concave curve shape at high compression ratios is influenced by the texture information and also by the motion information. Overall, the models result in good predictions of the actual curves and are based on the statistical parameters for texture and motion bits, estimated from only two sample encodings.

ACKNOWLEDGMENT
Supported in part by NSF Grant No. ANI-0136774.

4. REFERENCES