We are IntechOpen, the world’s leading publisher of Open Access books
Built by scientists, for scientists

4,000
Open access books available

116,000
International authors and editors

120M
Downloads

154
Countries delivered to

TOP 1%
Our authors are among the most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™
Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com
1. Introduction

MPEG standards family specify the decoding process and the bit-stream syntaxes allowing research towards the optimizations of the encoding process regarding coding performance improvement and complexity reduction. The purpose of a video encoder for broadcast or storage is to generate the optimal perceptual video quality, or the minimized distortion, under a certain constraint such as storage space or channel bandwidth. In particular, by minimizing the distortion $D$, the video encoder should optimally compute a set of optimal quantisers to control the output bit-rate for each coding unit to satisfy the allocated bit budget.

There are two main approaches to solve the optimal bit allocation problem: Lagrange optimization (Everett, 1963; Ramchandran et al., 1994) and dynamic programming (DP) (Bellman, 2003). The optimal bit allocation was first addressed in (Huang & Schultheiss, 1963) where the Lagrange multiplier approach for R-D analysis in transform coding was used. Further improvements have been reported in (Shoham & Gersho, 1988) for source quantization and coding. However, the Lagrange multiplier method suffer from problems, such as having negative bits and real numbers (Schuster & Katsaggelos, 1997a) and the computational complexity is very high due to the need to determine R-D characteristics of current and future video frames. DP is employed to achieve the minimum overall distortion through a tree or trellis with known quantisers and their R-D characteristics (Forney, 1973; Ortega, 1996; Ramchandran et al., 1994).

The total of the required bits and coding distortion depend on the quantization step-size. The rate or distortion versus quantization parameter (Q) curve can be produce by encoding for all the possible quantisers to obtain the bit-rate and the quantization error. In order to know how to select a quantization parameter under a specific constraint, e.g., the target bit-budget or distortion, it is importance to model or estimate the coding bit rate in terms of the quantization parameter, namely rate-quantization (R-Q) functions. Together with distortion-quantization (D-Q) functions, R-Q functions characterize the rate-distortion (R-D) behaviour of video encoding, which is the key to obtain an optimum bit allocation. Many R-Q and D-Q functions have been reported in previous studies (Chiang & Zhang, 1997; Ding & Liu, 1996; Hang & J.J. Chen, 1997; ISO/IEC, 1993; ISO/IEC, 1997; ITU-T, 1997; Lin & Ortega, 1998;
Ortega, 1996; Ribas-Corbera & Lei, 1999; Sullivan & Wiegand, 1998; Yin & Boyce, 2004). Some of these schemes were adopted in standard-compliant video coders, such as TM-5 (ISO/IEC, 1993), the test model for MPEG-2, TMN-8 (ITU-T, 1997), the test model for H.263, and VM-8 (ISO/IEC, 1997), the verification model for MPEG-4.

Usually rate control algorithms accept as an assumption that video source statistics are stationary. In this case, video source statistics correspond to some form of probability model such as Gaussian (Hang & J. J. Chen, 1997) or Laplacian (Chiang & Zhang, 1997) and R-D models based on the R-D theory, the theoretical foundation of rate control, can be obtained (Berger, 1971; Chiang & Zhang, 1997; Ribas-Corbera & Lei, 1999).

A video coding algorithm focus on the trade-off between the distortion and bit rate, where usually to a decreasing distortion corresponds an increasing rate and vice-versa. In R-D theory, the R-D function allows to estimate the lower bound for the rate at a given distortion. However, this value may not be possible to obtain in practical video encoders implementations. Operational R-D (ORD) theory applies to lossy data compression with finite number of possible R-D pairs (Schuster & Katsaggelos, 1997a).

Fig. 1. Operational rate-distortion and rate-distortion model curves.

The ORD function presents the convex curve of the specific compression scheme such that the optimal solution of rate control, i.e., optimal quantiser achieving minimum distortion at given bit rate, can be obtained (Schuster & Katsaggelos, 1997a) (Figure 1). Efficiency problems in many practical video coding applications may occur due to high computational complexity in this approach (Z. Chen & Ngan, 2007). Therefore, in numerous systems, model-based rate control schemes have been adopted (Chiang & Zhang, 1997; Ding & Liu, 1996; Vetro et al., 1999; Z. Chen & Ngan, 2005a; Zhang et al., 2005). R-D models can be obtained based on the statistical properties of video signal and R-D theory (Chiang & Zhang, 1997; Hang & J. J. Chen, 1997; Ribas-Corbera & Lei, 1999), or on empirical observation and benefiting from various regression techniques (Ding & Liu, 1996; Kim, 2003; Lin & Ortega, 1998; Z. Chen & Ngan, 2004; Z. Chen & Ngan, 2005b).
Some rate control schemes incorporate spatio-temporal correlations to improve the accuracy of R-D models, by using statistical regress analysis for dynamical model parameters update. Representative of this approach is the MPEG-4 Q2 (Chiang & Zhang, 1997), and the linear MAD models (Lee et al., 2000), where model parameters are updated by linear regression method from previous coded parameters. H.264/AVC JM rate-control algorithm also uses a quadratic rate model. In addition, the H.264/AVC rate-control solves “chicken-and-egg” dilemma as the Lagrange multiplier is modelled as a function of quantization parameter (Wiegand & Girod, 2001). Rate-quantization relationship can be used to compute the quantization parameter. Nevertheless, the model-based rate functions frequently depend on the complexity of the coding unit that is obtained after the rate-constrained motion estimation and mode decision with the Lagrange multiplier. The JM algorithm of H.264/AVC proposes a linear prediction model to solve this problem by estimating the mean of absolute difference (MAD) from the previously coded units. Then the quadratic model can estimate the quantization parameter. However, rate-distortion re-analysis can be further investigated based on the coding characteristics of the H.264/AVC for improving the coding performance (Kamaci et al., 2005; Ma et al., 2005) particularly in the case of joint video coding and the use of different distortion metrics.

We may find in the literature extensive studies regarding optimizing a video encoder with R-D considerations include mode decision (Chan & Siu, 2001; Chung & Chang, 2003), motion estimation (Pur et al., 1987; Rhee et al., 2001; Wiegand et al., 2003b), optimal bit allocation and rate control in video coding field (H.-Y.C. Tourapis & A.M. Tourapis, 2003; He & Mitra, 2002; J.J. Chen & Lin 1996; Ortega, 1996; Ramchandran et al., 1994; Ribas-Corbera & Neuhoff, 1998; Schuster & Katsaggelos, 1997b; Sullivan & Wiegand, 1997; Wiegand et al., 2003a, 2003c; Zhang et al., 2003).

In summary, to optimize a video encoder, the rate-distortion optimization techniques play a very important role. R-D models are functions that predict the expected distortion at a given bit rate. This is very important for joint video coding applications that attempt to optimized quality, e.g. minimize distortion, in environments where the channel conditions vary dynamically or the number of broadcast programs varies through time. Thus in this section we propose to present and evaluate several R-D models.

At the same time, we propose also to study the bit rate variability as a function of the video quality (Seeling et al., 2004, 2007). This type of analyse is typical of a communication network perspective. By re-analyzing the characteristics of the bit-rate and the data in the transform domain, a simple rate estimation function can be obtained that will allow support the allocating of video bandwidth within different video programmes.

2. Rate control in international standards

Although the MPEG video coding standard recommended a general coding methodology and syntax for the creation of a legitimate MPEG bitstream, there are many areas of research left open regarding how to generate high-quality MPEG bitstreams. This allows the designers of MPEG encoder great flexibility in developing and implementing their own MPEG specific algorithms. To optimise the performance-of an MPEG encoder system, it is important to study research areas such as motion estimation, coding mode decisions, and rate control.

The main goal of rate control is to manage the process of bit allocation within a video sequence and thus the quality of the encoded bitstream. Regarding rate control, encoders
can operate at Constant Bit Rate (CBR) or in Variable Bit Rate (VBR). In CBR, the video encoder maintains the average bit rate constant. The encoder output has a buffer and its occupancy is controlled dynamically by adjusting the quantization scale, denoted as \( q \) in MPEG coders. Likewise, the quality of the video sequence varies due to the variations in the scene complexity. VBR reduces the variation in the picture quality by allocating more bits to complex images. A common use of VBR is Open-Loop Variable Bit Rate (OL-VBR), where the quantization scale is constant for all the images of the video sequence. Another VBR scheme is Constant Quality - Variable Bit Rate (CQ-VBR) which aims to maintain an objective video quality constant.

The rate control algorithms usually adjust the coded bit stream according to different constraints, such as buffer over- or underflow prevention, variable and/or low bandwidth constraints resulting from limited storage size or communication bandwidth (Ortega, 1996). In order to accomplish this goal, rate control schemes are responsible to adjust the quantization parameters.

A generic bit rate control is composed of the following steps: given an input video signal and a desired bit rate, constant or variable, what should be the encoder settings to maintain the picture quality as high and constant as possible. In MPEG encoding, a quantization scale controls the trade-off between picture quality and the bit rate. This parameter is used to compute the step size of the uniform quantisers used for the different AC DCT coefficients (ISO/IEC, 1993). For each macroblock, a quantiser, \( q \), is selected. It is named “adaptive quantization” to the process for adjusting the value of \( q \) between macroblocks within an image frame. There are several schemes for doing the adaptive quantization. For example, in MPEG-2 Test Model 5 (TM5) (ISO/IEC, 1993), a non-linear mapping based on the block variance is used to adapt the \( q \)’s. Besides the quantization scale, the quantization coarseness is also dependent on the quantization matrix. In MPEG-1, the quantization matrix can be altered in each sequence while in MPEG-2 on a picture basis. It sets the relative coarseness of quantization for each coefficient.

As MPEG does not specify how to control the bit rate, different approaches can be found in the literature (ISO/IEC, 1993; Keesman et al., 1995; Ramchandran et al., 1993). Two approaches have been used: ‘feed forward bit rate control’ and ‘feed backward bit rate control’. In the first approach, after performing a pre-analysis, the optimum settings are compute. This process will increase the computational complexity and time needed while yielding better results. In the second approach, there is limited knowledge of the sequence complexity. Bits are allocated on a picture basis and spatially uniform distributed throughout the image. Thus, too many bits may be spent at the beginning of the picture while the end of the picture may present a higher degree of complexity. The ‘feed backward bit rate control’ is suitable for real-time applications and ‘feed forward bit rate control’ for applications where the quality is the main goal and time is not a constraint.

---

Fig. 2. Rate control in video coding system.

---

www.intechopen.com
3. Rate control in H.264/AVC

Existing studies indicate that H.264/AVC brings major improvement in coding performance in relation to prior coding standards (Wiegand et al., 2003a). H.264/AVC presents many new features, which represent huge challenges to the operative encoder control such as how to allocate the bandwidth between the texture coding and the overhead coding.

A major contributor to the high coding efficiency of H.264/AVC compared with previous video compression standards is the rate and distortion (R-D) optimized motion estimation and mode decision (also referred to as RDO) with various intra and inter prediction modes and multiple reference frames. Nevertheless, these innovations increase the rate control process complexity due to the inter-dependency between the RDO and rate control. Only after the end of intra/inters prediction, the rate control scheme can access the exact coding characteristics. This information is necessary for the computation of the quantization parameter. Such a dilemma prevents the rate control scheme from directly accessing the coding characteristic in advance. The dilemma of selecting which parameter should be first determine is sometimes referred in the literature as to the “chicken and egg” dilemma (Li et al., 2003c, 2004; Wu et al., 2005).

To avoid this dilemma, in JVT-D030 a two-pass scheme was proposed, where in each pass a TM-5-alike method was used (Ma et al., 2002). This approach uses an extremely simplified R-D function, which fails to achieve accurate and robust rate control and due to the two-pass increase the level of complexity. Because of these drawbacks, JVT-G012 (Li et al., 2003a) was proposed and accepted as the standardized rate control scheme for H.264/AVC. In JVT-G012, a linear MAD model predicts the coding complexity, and a MPEG-4 Q2 function employed to estimate the quantization parameter (Li et al., 2003a).

First step occurs at GOP level. This step estimates the bits available for the remaining frames in the GOP. In addition, it initializes the QP of instantaneous decoding refresh (IDR) frame. In the following step, rate control algorithm operates at Picture level: an estimation of the target bits for the current basic unit is determined. A basic unit is a group of macroblocks and its size can vary from one macroblock up to the entire picture. The target bits estimation should be allocate so that a similar number of bits are allocate for every picture and the target buffer level is preserve.

The next step is, based on the number of bits used to encode the previous basic units, to estimate the necessary bits to encode the header. The target texture is obtained by subtracting the header estimation to the total target bits estimate. After that, this value is converted to a target QP value using a quadratic model that correlates the QP with the texture bits. The quadratic model needs an estimation of the MAD of the motion-compensated or intra prediction error of the current basic unit’s. Consequently, the rate control model requires an additional linear MAD model that, from the previous basic unit MAD, allows the computaion of the current basic unit MAD. In summary, the Picture level process consist in computing the quantization step Qstep using a quadratic model and then performing a R-D optimization (RDO) (Wiegand et al., 2003a) for each MB in the frame.

The MAD of the current stored picture, , is predicted by a linear regression model similar to that of MPEG-4 Q2 after coding each picture or each basic unit (1) using the actual MAD of the previous stored picture, \(\sigma_i(j - 1 - L)\)

\[
\tilde{\sigma}_i(j) = a_i \times \sigma_i(j - 1 - L) + a_z
\]  

(1)
where $a_1$ and $a_2$ are the model parameters (first-order and second-order coefficients). The initial value of $a_1$ and $a_2$ are set to one and zero, respectively (Lim et al., 2007). The quantization step corresponding to the target bits is computed by the equation (2)

\[ T_i(j) = c_1 \times \frac{\tilde{\sigma}_i(j)}{Q_{\text{step},i}(j)} + c_2 \times \frac{\tilde{\sigma}_i^2(j)}{Q_{\text{step},i}^2(j)} - m_{h_i}(j) \]

where $m_{h_i}(j)$ is the total number of header bits and motion vector bits, $c_1$ and $c_2$ are two coefficients. The corresponding quantization parameter $Q_{\text{Pi}}(j)$ is computed by using the relationship between the quantization step and the quantization parameter of AVC (Lim et al., 2007). Final step consists in updating the quadratic QP/bits linear model and the MAD model. This process repeats for each basic unit until the complete video sequence has been encoded.

In this section, it was introduce the basis of rate control architecture in JM H.264/AVC (Lim et al., 2007). More detail information is available at (Li et al., 2003b, 2003c, 2004; Lim et al., 2007; Ma et al., 2002). Other solutions can be found in the literature. For example, Zhihai He (He, 2001) has proposed a new model that achieves a good performance for H.263 and MPEG4-2 (ISO/IEC 14496-2) codecs. The parameter $\rho$ represents the percentage of zeros among the quantized transform coefficients. He found a linear relationship between the value $\rho$ and the real bit rate because the percentage of zeros plays an important role in determining the final bit rate.

4. Test video sequences

Selecting a representative set of video sequences is a crucial step in evaluating and analysing the performance of R-D models. A homogeneous set of video sequences may generate biased comparison results, because some models may perform especially well under certain sequences. Two key features are used to characterize video sequences: spatial complexity and temporal complexity. Usually, spatial complexity is measured by averaging all neighbourhood differences in the same frame while temporal complexity is measured by averaging neighbourhood differences between adjacent frames (Adjeroh & Lee, 2004). The set of test video sequences is composed by twelve CIF video sequences, with the duration of 10 seconds that are known as test video sequences (ITU-T, 2005). It were included sequences with low spatial and temporal complexity (low complexity sequences) up to sequences with high spatial and temporal complexity (high complexity sequences). Sequences that have either high spatial or temporal complexity but nor both the designated them as medium-complexity sequences. It follows a brief description of the sequences.

In seven video sequences, the position of the camera is fixed: Akiyo (aki), Deadline (dea), Hall (hal), Mother and Daughter (mad), News (new), Paris (par) and Silence (sil). In the Akiyo sequence, the camera is focus on a human subject with a synthetic background (a female anchor reading the news). The movements are very limited, mainly head movements in front of a fixed camera. In Deadline, Mother and Daughter and Paris sequences, the camera is still fixed but there are more movements of the bodies and heads. These are typical videoconferencing content. In the News sequence two reporters, a male and a female anchor, reading the news in front of a fixed camera in a newsroom while in the background, two dancers execute movements. Hall sequence is an example of a video supervision, with stationary camera and two moving persons: one people entering from the left with a
briefcase and then leaving the hall. In the middle of the sequence, a second person enters the hall from the right and then grabs a monitor. In the Silence sequence, one can observe a fast moving subject executing deaf gesture language.

![Video test sequences](image)

Fig. 3. Video test sequences

The Foreman sequence (for) contain the head of a person talking and geometric shapes. Fast camera movement and content motion with a pan to a construction site at the end characterize this sequence. The main characteristics of the Flower Garden sequence (flg) is the slow and steady camera panning over landscape over landscape; the spatial and the colour detail. Coastguard sequence (cgd) was shot as a pan from left to right movement in the first third and a pan from left to right in the rest of the sequence. The camera movement follows the movements of two boats (the first from right to left and the second movement from left to right). The Mobile and Calendar (mcl) sequence is characterized by the slow panning and zooming of the camera, complex motion; high spatial and colour detail. Fast and complex motion movements of the camera and contents and the level of detail characterize the Football sequence (fot). This is a very diverse set of video sequences

5. Experimental setup

Simulations were performed with the JM reference software, the official MPEG and ITU reference implementation, for the H.264/AVC Main profile (ITU-T, 2005). Source code was compiled with Microsoft Visual C++:

www.intechopen.com
Four different type of GOP patterns were used (Table 1). A typical GOP pattern (IBBP_GOP2), an "extend" B frame version of the typical GOP pattern, and two GOP patterns without Interpolated images.

Additionally, each video test sequence was encoded in two modes: Open-loop (fixed QP with values ranging from 10 up to 42) and Constant Bit Rate (Fixed Rate - 64kbps, 128kbps, 256kbps, 384kbps, 512kbps, 640kbps, 768kbps, 1024kbps, 1536kbps, 2048kbps. The goal was to obtain sufficient data to obtain R-D curves.

Typical quality metrics include Peak signal-to-noise ratio (PSNR) and the Mean Square Error (MSE), Sum of Squared Differences (SSD), Mean Absolute Difference (MAD), and Sum of Absolute Differences (SAD).

$$PSNR = 10 \log_{10} \frac{255^2}{MSE}$$  \hspace{1cm} (3)

$$MSE = \frac{1}{HW} SSD$$ \hspace{1cm} (4)

$$SSD = \sum_{x=0}^{H-1} \sum_{y=0}^{W-1} (p(x,y) - \hat{p}(x,y))^2$$ \hspace{1cm} (5)

$$MAD = \frac{1}{HW} SAD$$ \hspace{1cm} (6)

$$SAD = \sum_{x=0}^{H-1} \sum_{y=0}^{W-1} |p(x,y) - \hat{p}(x,y)|$$ \hspace{1cm} (7)

where \(H\) and \(W\) are the height and the width of the image frame, and \(p(x,y)\) and \(\hat{p}(x,y)\) represent the “original” and the reconstructed image frame pixels at \((x, y)\).

Complementary with the study of rate-distortion performance it is propose to include an analysis of the bit rate variability as a function of the video quality. This is an important topic when considering multimedia traffic. The bit rate variability is usually characterized by the Coefficient of Variation (CoV) of the frame sizes (in bits), whereby the CoV is defined as the standard deviation of the frame sizes normalized by their mean \(\bar{X}\) (Seeling et al., 2004, 2007)

$$CoV = \frac{\sigma}{\bar{X}}$$ \hspace{1cm} (8)

where \(\bar{X}\) is mean size (in bits)
and the variance $\sigma^2$ (square of the standard deviation) of the frame sizes being defined as

$$\sigma^2 = \frac{1}{(M-1)} \sum_{m=1}^{M} (X_m - \bar{X})^2$$

(10)

6. Experimental results and discussion

This section presents experimental results: Rate-Distortion analysis and bit rate variability analysis as a function of the video quality.

6.1 R-D models

The RD graphs obtained for the video sequences Akiyo, Foreman and Football, in open loop, are show in Figure 4 (bit-rate axe is in logarithm scale). One can observe that a proportional relation exists between Bit-rate and Picture Quality and that quality depend on the video nature: for the same bit-rate, low complexity sequences present higher values of quality and vice-versa. This behaviour occurs in all the different GOP patterns. Figure 5

![Fig. 4. Rate-distortion curve (Akiyo, Foreman, Football; FixeQP)](www.intechopen.com)
Recent Advances on Video Coding

126

present graphic representation for RD data in Constant Bit Rate for the same three video sequences using JM rate control. In this case a relation between bit rate and quality can be observed.

Fig. 5. Rate-distortion curve (Akiyo, Foreman, Football; FixeRate)

Frequently data can be noisy in its nature. Thus recognizing the trends in the data is important (Vardeman, 1994). One of the available methods for data analysis and identify existing trends in physical systems is curve fitting. The concept of curve fitting is rather simple: to use a simple function to describe a trend by minimizing the error between the selected function to fit and a set of data (Vardeman, 1994). The principle of least squares is applied to the fitting of a line to (x, y) data. Representative work for estimate the quantization step size has been most direct towards developing all kinds of rate-quantization (R-Q) models like polynomial (including linear and quadratic) (Chiang & Zhang, 1997; Lin & Ortega, 1998; Ronda et al., 1999; Yan & Liou, 1997), spline (Lin et al., 1996), logarithmic (Ding & Liu, 1996; Hang & J.J. Chen, 1997), power (Ding & Liu, 1996), etc. Yang et al. (Kyeong Ho Yang et al, 1997) proposed a more complex model that combines a logarithmic and a quadratic model. Most of the models only consider the rate function, and often implicitly assume that the distortion is a linear function of the quantization scale. This work has been extended to include D(QP) implementing several methods in order to compare their results. In fact, the goal is to model the quality versus quantization step

www.intechopen.com
relationship and then to evaluate the different approaches to quality metric. It is presume that there is an inverse relationship between quality and distortion. Before fitting data into a function that models the relationship between two measured quantities, it is a normal procedure to determine if a relationship exists between these quantities. It was decide to use the correlation method to confirm the degree of probability that a relationship exists between two measured quantities (Vardeman, 1994). In the case of no correlation between the two quantities, then there is no tendency for the values of one quantity to increase or decrease with the values of the second quantity. To evaluate the quality of the fit, it is used the sample correlation that represents the normalized measure of the strength of linear relationship between variables (Vardeman, 1994):

\[
 r = \frac{x'y}{\sqrt{(x'x)(y'y)}} \tag{11}
\]

\[
 r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} = \frac{\sum x_i y_i - \left(\frac{\sum x_i}{n}\right)\left(\frac{\sum y_i}{n}\right)}{\sqrt{\left[\sum x_i^2 - \left(\frac{\sum x_i}{n}\right)^2\right] \left[\sum y_i^2 - \left(\frac{\sum y_i}{n}\right)^2\right]}} \tag{12}
\]

where \( r \) is a matrix of correlation coefficients (Vardeman, 1994). The sample correlation always lies in the interval from -1 to 1. A value of \( r \) near of positive one or negative one, it is interpreted as indicating a relatively strong relationship and \( r \) near zero is inferred as indicating a lack of relationship. The sign of \( r \) indicates whether \( y \) tends to increase or decrease with increase \( x \).

<table>
<thead>
<tr>
<th>Sequence</th>
<th>I Type</th>
<th>B Type</th>
<th>I Type</th>
<th>B Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aki</td>
<td>0.8380</td>
<td>0.9016</td>
<td>0.8645</td>
<td>0.9034</td>
</tr>
<tr>
<td>Cgd</td>
<td>0.9210</td>
<td>0.9595</td>
<td>0.9180</td>
<td>0.9609</td>
</tr>
<tr>
<td>Dea</td>
<td>0.8853</td>
<td>0.9303</td>
<td>0.8943</td>
<td>0.9609</td>
</tr>
<tr>
<td>Flg</td>
<td>0.9137</td>
<td>0.9349</td>
<td>0.9147</td>
<td>0.9342</td>
</tr>
<tr>
<td>For</td>
<td>0.8964</td>
<td>0.9035</td>
<td>0.8968</td>
<td>0.9321</td>
</tr>
<tr>
<td>Fot</td>
<td>0.9588</td>
<td>0.9557</td>
<td>0.9657</td>
<td>0.9695</td>
</tr>
<tr>
<td>Hal</td>
<td>0.8154</td>
<td>0.7972</td>
<td>0.8030</td>
<td>0.8628</td>
</tr>
<tr>
<td>Mad</td>
<td>0.8797</td>
<td>0.8666</td>
<td>0.8653</td>
<td>0.9129</td>
</tr>
<tr>
<td>New</td>
<td>0.9554</td>
<td>0.9081</td>
<td>0.9511</td>
<td>0.9524</td>
</tr>
<tr>
<td>Par</td>
<td>0.9455</td>
<td>0.9435</td>
<td>0.9461</td>
<td>0.9651</td>
</tr>
<tr>
<td>Sil</td>
<td>0.9451</td>
<td>0.9412</td>
<td>0.9419</td>
<td>0.9576</td>
</tr>
<tr>
<td>Mcl</td>
<td>0.9356</td>
<td>0.9272</td>
<td>0.9329</td>
<td>0.9488</td>
</tr>
</tbody>
</table>

Table 2. Correlation coefficients between Bits Frames and Quality Metric (PSNR) for different H.264/AVC video sequences (IBBP-GOP1 and IBBP-GOP2).
Table 3. Correlation coefficients between Bits Frames and Quality Metric (PSNR) for different H.264/AVC video sequences (IPPP-GOP1 and IPPP-GOP2).

<table>
<thead>
<tr>
<th>Sequence</th>
<th>IPPP – GOP1</th>
<th>IPPP – GOP2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aki</td>
<td>0.8962</td>
<td>0.9170</td>
</tr>
<tr>
<td>Cgd</td>
<td>0.9608</td>
<td>0.9617</td>
</tr>
<tr>
<td>Dea</td>
<td>0.9304</td>
<td>0.9406</td>
</tr>
<tr>
<td>Flg</td>
<td>0.9555</td>
<td>0.9586</td>
</tr>
<tr>
<td>For</td>
<td>0.9268</td>
<td>0.9333</td>
</tr>
<tr>
<td>Fot</td>
<td>0.9659</td>
<td>0.9686</td>
</tr>
<tr>
<td>Hal</td>
<td>0.8540</td>
<td>0.8848</td>
</tr>
<tr>
<td>Mad</td>
<td>0.9019</td>
<td>0.9200</td>
</tr>
<tr>
<td>New</td>
<td>0.9353</td>
<td>0.9492</td>
</tr>
<tr>
<td>Par</td>
<td>0.9609</td>
<td>0.9668</td>
</tr>
<tr>
<td>Sil</td>
<td>0.9605</td>
<td>0.9662</td>
</tr>
<tr>
<td>Mcl</td>
<td>0.9584</td>
<td>0.9715</td>
</tr>
</tbody>
</table>

Equation (12) was computed for all the twelve sequences, and results were obtained according the different Picture Type and GOP pattern (Table 2 and Table 3). Thus, it was assess the hypothesis of a relationship between PSNR and Rate. Results are very high, for all the video sequences and GOP patterns, near positive one, pointing clearly to a strong positive linear relationship evident. Next step is thus to select what curve fitting functions should be assessed. Due to its simplicity, the first selected is one of the most commonly used techniques: the fitting of a straight line to a set of bivariate data generating a linear equation such as (13) (Vardeman, 1994):

Linear \[ y = \beta_0 + \beta_1x \] (13)

A natural generalization of equation (13) is the polynomial equation (14)

Polynomial \[ y = \beta_0 + \beta_1x + \beta_2x^2 + \ldots + \beta_kx^k \] (14)

The goal is thus to minimize the function of \( k + 1 \) variables.

\[
S(\beta_0, \beta_1, \beta_2, \ldots, \beta_k) = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2
= \sum_{i=1}^{n}(y_i - (\beta_0 + \beta_1x_i + \beta_2x_i^2 + \ldots + \beta_kx_i^k))^2
\] (15)

by selecting the coefficients \( \beta_0, \beta_1, \beta_2, \ldots, \beta_k \) (Vardeman, 1994). Upon setting the partial derivatives of \( S(\beta_0, \beta_1, \beta_2, \ldots, \beta_k) \) equal to zero and doing some simplifications, one obtains the normal equations for this least squares problem:

\[
\begin{align*}
n\beta_0 + (\sum x_i)\beta_1 + (\sum x_i^2)\beta_2 + \ldots + (\sum x_i^k)\beta_k &= \sum y_i \\
(\sum x_i)\beta_0 + (\sum x_i^2)\beta_1 + (\sum x_i^3)\beta_2 + \ldots + (\sum x_i^{k+1})\beta_k &= \sum x_iy_i \\
(\sum x_i^2)\beta_0 + (\sum x_i^{k+1})\beta_1 + (\sum x_i^{k+2})\beta_2 + \ldots + (\sum x_i^{2k})\beta_k &= \sum x_i^2y_i
\end{align*}
\] (16)
Solving the system of $k+1$ linear equations presented in Equation 16 it is typically possible to obtain a single set of values $S(b_0, b_1, b_2, ..., b_k)$ that minimize $S(b_0, b_1, b_2, ..., b_k)$. Polynomials are often used when a simple empirical model is required. One of the most uses polynomial models is the quadratic model (Equation 17):

$$\text{Quadratic} \quad y = \beta_0 + \beta_1 x + \beta_2 x^2$$ (17)

To compare with the solution available in the literature it was decide to extend the models and thus include the logarithmic (18), the exponential (19), the power (20) and the linear with nonpolynomial model (LNP) (21).

$$\text{Logarithmic} \quad y = \beta_0 + \log x$$ (18)

$$\text{Exponential} \quad y = \beta_0 e^{\beta_1 x}$$ (19)

$$\text{Power} \quad y = \beta_0 + \beta_1 x^{\beta_2}$$ (20)

$$\text{Linear with nonpolynomial} \quad y = \beta_0 + \beta_1 e^{-x} + \beta_2 x e^{-x}$$ (21)

After selecting these six models, it was computed the average absolute error when trying to model the relation between bit-rate and quantization parameter (QP), PSNR and quantization parameter, and bit-rate and PSNR regarding the picture type using each of the six models for all the GOP patterns.

```
1. for each method do
  2. square error R(QP)(Picture Type) = 0;
  3. square error D(QP)(Picture Type) = 0;
  4. for each frame in the sequence do
    5. for each QP do
      6. Extract Statistics [Bits, PSNR, Picture Type];
      7. endfor
    8. Estimate the parameters of the model for R(QP) (Picture Type);
    9. Compute the square error R for each D value (Picture Type);
   10. Update the accumulative squared error R(Picture Type);
   11. Estimate the parameters of the model for D(QP) (Picture Type);
   12. Compute the square error D for each D value (Picture Type);
   13. Update the accumulative squared error D(Picture Type);
   14. Estimate the parameters of the model for R(D) (Picture Type);
   15. Compute the square error R for each D value(Picture Type);
   16. Update the accumulative squared error R_D(Picture Type);
  17. endfor
  18. endfor
```

Fig. 6. Pseudo code for R-D model fitting.

It was implemented the procedure described in Figure 6. Results are presented in Table 4, Table 5, and Table 6 for the twelve video sequences.
From the results, several observations can be produce. First, the linear with nonpolynomial model is the least accurate while the quadratic approach is the most accurate overall. The second observation is that the accuracy of all models varies with the level of complexity of the video source data. Results improve for low complexity video sequences while decrease for sequence with higher complexity. Third observation, GOP pattern has impact on the average of the absolute error for the different type of pictures. For most of the models, the average absolute error (excluding linear with nonpolynomial model) is rather small.
Considering individual video sequence results, they can be analysed according model fit, picture type, and GOP pattern for the different rate-distortion-quantization models. Regarding Rate-QP, quadratic approach is the best solution in most of the cases (for IPPP GOP1 and IPPP GOP2 quadratic approach is the best solution for 9 video sequences regarding picture type Intra and 10 video sequences for pictures type P and for the remaining video sequences the best solution is the exponential fit and power regression). Worst results of quadratic approach take place with IBBP GOP1 and IBBP GOP2 patterns (regarding picture type I, P and B, quadratic approach present the best results in 11, 6 and 8 video sequences for IBBP GOP1 and 10, 6 and 10 for IBBP GOP2). Besides quadratic approach, exponential fit and power regression also present good results, particularly in GOP patterns containing B images and for low to medium spatial and temporal complexity where motion estimation is most effective. In these cases, quadratic approach is usually the second best approach. Finally, quadratic is also the best approach for modelling Rate-PSNR (11 of 12 video sequences for IPPP GOP1 for both I and P frame types; 10 and 11 of 12 video sequences regarding respectively Intra and P frames for IPPP GOP2; 10, 11 and 10 for I, P and B frames regarding IBBP GOP1 and 10, 9 and 10 for I, P and B frames regarding IBBP GOP2). In this case, also exponential and power regression presents good results. Thus, quadratic approach is a good solution particularly for GOP sequences without B frames. For quality versus quantization parameter global results from different models are very good. In
### Recent Advances on Video Coding

Table 8. Absolute error for Rate-QP, PSNR-QP and Rate-PSNR curve fitting (IPPP GOP2).

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Fit Method</th>
<th>Rate-QP</th>
<th>PSNR-QP</th>
<th>Rate - PSNR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>I Type</td>
<td>P Type</td>
<td>B Type</td>
</tr>
<tr>
<td>Akiyo</td>
<td>Linear fit</td>
<td>462</td>
<td>228</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>Quadratic fit</td>
<td>108</td>
<td>43</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>Exponential fit</td>
<td>111</td>
<td>39</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>Logarithmic fit</td>
<td>339</td>
<td>157</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>Power Regression</td>
<td>212</td>
<td>112</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>LNP fit</td>
<td>791</td>
<td>451</td>
<td>1.18</td>
</tr>
<tr>
<td>Foreman</td>
<td>Linear fit</td>
<td>1776</td>
<td>717</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>Quadratic fit</td>
<td>460</td>
<td>169</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>Exponential fit</td>
<td>434</td>
<td>277</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>Logarithmic fit</td>
<td>1444</td>
<td>550</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>Power Regression</td>
<td>542</td>
<td>450</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>LNP fit</td>
<td>1707</td>
<td>1045</td>
<td>1.11</td>
</tr>
<tr>
<td>Football</td>
<td>Linear fit</td>
<td>3043</td>
<td>1068</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>Quadratic fit</td>
<td>532</td>
<td>179</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>Exponential fit</td>
<td>664</td>
<td>250</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>Logarithmic fit</td>
<td>2212</td>
<td>774</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>Power Regression</td>
<td>1974</td>
<td>711</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>LNP fit</td>
<td>4872</td>
<td>1722</td>
<td>1.09</td>
</tr>
</tbody>
</table>

Table 9. Absolute error for Rate-QP, PSNR-QP and Rate-PSNR curve fitting (IBBP GOP1).

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Fit Method</th>
<th>Rate-QP</th>
<th>PSNR-QP</th>
<th>Rate - PSNR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>I Type</td>
<td>P Type</td>
<td>B Type</td>
</tr>
<tr>
<td>Akiyo</td>
<td>Linear fit</td>
<td>1063</td>
<td>115</td>
<td>221</td>
</tr>
<tr>
<td></td>
<td>Quadratic fit</td>
<td>190</td>
<td>46</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>Exponential fit</td>
<td>218</td>
<td>56</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>Logarithmic fit</td>
<td>707</td>
<td>99</td>
<td>163</td>
</tr>
<tr>
<td></td>
<td>Power Regression</td>
<td>605</td>
<td>37</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>LNP fit</td>
<td>2269</td>
<td>65</td>
<td>368</td>
</tr>
<tr>
<td>Foreman</td>
<td>Linear fit</td>
<td>2736</td>
<td>726</td>
<td>769</td>
</tr>
<tr>
<td></td>
<td>Quadratic fit</td>
<td>894</td>
<td>312</td>
<td>209</td>
</tr>
<tr>
<td></td>
<td>Exponential fit</td>
<td>1123</td>
<td>389</td>
<td>286</td>
</tr>
<tr>
<td></td>
<td>Logarithmic fit</td>
<td>2175</td>
<td>634</td>
<td>605</td>
</tr>
<tr>
<td></td>
<td>Power Regression</td>
<td>1132</td>
<td>265</td>
<td>428</td>
</tr>
<tr>
<td></td>
<td>LNP fit</td>
<td>3396</td>
<td>414</td>
<td>970</td>
</tr>
<tr>
<td>Football</td>
<td>Linear fit</td>
<td>5893</td>
<td>1564</td>
<td>1368</td>
</tr>
<tr>
<td></td>
<td>Quadratic fit</td>
<td>1038</td>
<td>289</td>
<td>241</td>
</tr>
<tr>
<td></td>
<td>Exponential fit</td>
<td>1779</td>
<td>558</td>
<td>369</td>
</tr>
<tr>
<td></td>
<td>Logarithmic fit</td>
<td>4261</td>
<td>1179</td>
<td>991</td>
</tr>
<tr>
<td></td>
<td>Power Regression</td>
<td>4411</td>
<td>1254</td>
<td>962</td>
</tr>
<tr>
<td></td>
<td>LNP fit</td>
<td>9912</td>
<td>2162</td>
<td>2207</td>
</tr>
</tbody>
</table>

www.intechopen.com
this case, linear fit results are very interesting as although they are not among the best approaches, the error is rather small, particularly for low complex video sequences. These results indicate that aggregate video results might be represented by the following equations:

\[ R = \beta_1 + \beta_2 QP + \beta_3 QP^2 \]  \hspace{1cm} (22)

\[ PSNR = \beta'_1 + \beta'_2 \times QP + \beta'_3 \times QP^2 \]  \hspace{1cm} (23)

\[ R = \beta''_1 + \beta''_2 \times PSNR + \beta''_3 \times PSNR^2 \]  \hspace{1cm} (24)

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Fit Method</th>
<th>Rate-QP I</th>
<th>Rate-QP P</th>
<th>Rate-QP B</th>
<th>PSNR-QP I</th>
<th>PSNR-QP P</th>
<th>PSNR-QP B</th>
<th>Rate-PSNR I</th>
<th>Rate-PSNR P</th>
<th>Rate-PSNR B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Akiyo</td>
<td>Linear fit</td>
<td>467</td>
<td>120</td>
<td>296</td>
<td>0.04</td>
<td>0.05</td>
<td>0.02</td>
<td>2452</td>
<td>1056</td>
<td>4331</td>
</tr>
<tr>
<td></td>
<td>Quadratic fit</td>
<td>109</td>
<td>49</td>
<td>58</td>
<td>0.03</td>
<td>0.02</td>
<td>0.01</td>
<td>644</td>
<td>447</td>
<td>939</td>
</tr>
<tr>
<td></td>
<td>Exponential fit</td>
<td>118</td>
<td>59</td>
<td>48</td>
<td>0.12</td>
<td>0.07</td>
<td>0.04</td>
<td>606</td>
<td>547</td>
<td>740</td>
</tr>
<tr>
<td></td>
<td>Logarithmic fit</td>
<td>328</td>
<td>104</td>
<td>206</td>
<td>0.19</td>
<td>0.12</td>
<td>0.07</td>
<td>2835</td>
<td>1124</td>
<td>5009</td>
</tr>
<tr>
<td></td>
<td>Power Regression</td>
<td>247</td>
<td>38</td>
<td>142</td>
<td>0.27</td>
<td>0.16</td>
<td>0.1</td>
<td>897</td>
<td>664</td>
<td>1358</td>
</tr>
<tr>
<td></td>
<td>LNP fit</td>
<td>915</td>
<td>67</td>
<td>568</td>
<td>1.29</td>
<td>0.75</td>
<td>0.46</td>
<td>8218</td>
<td>1961</td>
<td>14468</td>
</tr>
<tr>
<td>Foreman</td>
<td>Linear fit</td>
<td>1559</td>
<td>768</td>
<td>861</td>
<td>0.08</td>
<td>0.04</td>
<td>0.03</td>
<td>7387</td>
<td>6307</td>
<td>10540</td>
</tr>
<tr>
<td></td>
<td>Quadratic fit</td>
<td>565</td>
<td>331</td>
<td>206</td>
<td>0.04</td>
<td>0.02</td>
<td>0.01</td>
<td>2144</td>
<td>2255</td>
<td>2173</td>
</tr>
<tr>
<td></td>
<td>Exponential fit</td>
<td>677</td>
<td>430</td>
<td>342</td>
<td>0.09</td>
<td>0.05</td>
<td>0.03</td>
<td>2361</td>
<td>2753</td>
<td>6879</td>
</tr>
<tr>
<td></td>
<td>Logarithmic fit</td>
<td>1297</td>
<td>671</td>
<td>652</td>
<td>0.15</td>
<td>0.09</td>
<td>0.05</td>
<td>8159</td>
<td>6801</td>
<td>12310</td>
</tr>
<tr>
<td></td>
<td>Power Regression</td>
<td>544</td>
<td>296</td>
<td>575</td>
<td>0.23</td>
<td>0.14</td>
<td>0.08</td>
<td>3318</td>
<td>3672</td>
<td>5214</td>
</tr>
<tr>
<td></td>
<td>LNP fit</td>
<td>1517</td>
<td>431</td>
<td>1315</td>
<td>1.24</td>
<td>0.72</td>
<td>0.44</td>
<td>17604</td>
<td>11880</td>
<td>31947</td>
</tr>
<tr>
<td>Football</td>
<td>Linear fit</td>
<td>3228</td>
<td>1714</td>
<td>1428</td>
<td>0.12</td>
<td>0.07</td>
<td>0.05</td>
<td>13308</td>
<td>12846</td>
<td>15837</td>
</tr>
<tr>
<td></td>
<td>Quadratic fit</td>
<td>547</td>
<td>328</td>
<td>237</td>
<td>0.04</td>
<td>0.03</td>
<td>0.02</td>
<td>2263</td>
<td>2280</td>
<td>2469</td>
</tr>
<tr>
<td></td>
<td>Exponential fit</td>
<td>1047</td>
<td>609</td>
<td>398</td>
<td>0.08</td>
<td>0.05</td>
<td>0.03</td>
<td>10027</td>
<td>8032</td>
<td>12944</td>
</tr>
<tr>
<td></td>
<td>Logarithmic fit</td>
<td>2368</td>
<td>1295</td>
<td>1024</td>
<td>0.11</td>
<td>0.06</td>
<td>0.04</td>
<td>16052</td>
<td>15114</td>
<td>19319</td>
</tr>
<tr>
<td></td>
<td>Power Regression</td>
<td>2510</td>
<td>1357</td>
<td>1031</td>
<td>0.20</td>
<td>0.11</td>
<td>0.06</td>
<td>5714</td>
<td>4359</td>
<td>7924</td>
</tr>
<tr>
<td></td>
<td>LNP fit</td>
<td>5049</td>
<td>2352</td>
<td>2398</td>
<td>1.27</td>
<td>0.73</td>
<td>0.44</td>
<td>38424</td>
<td>33451</td>
<td>51582</td>
</tr>
</tbody>
</table>

Table 10. Absolute error for Rate-QP, PSNR-QP and Rate-PSNR curve fitting (IBBP GOP2)

7. Rate variability as a function of the video quality.

A second important issue for joint video coding broadcasting is the Rate Variability-Distortion (VD). Two sub-sets have been consider from the initial set of twelve video sequences: a first sub-set with camera movement, medium to high spatial detail and temporal complexity (sequences Foreman, Football, Coastguard, Flower Garden, and Mobile and Calendar), and a second sub-set with fixed camera and low to medium spatial detail and motion activity (Akiyo, Deadline, Hall, Mother and Daughter, News, Paris, and Silence). Results are presented in Figure 7, Figure 8, Figure 9, and Figure 10. In the left side it can be observe the results from the first sub-set and in the right the charts for the second sub-set. Simulations results are from open-loop coding setup.
Fig. 7. Rate Variability-distortion (VD) Curve (PSNR; IBBP GOP1).

Fig. 8. Rate Variability-distortion (VD) Curve (PSNR; IBBP GOP2).

Fig. 9. Rate Variability-distortion (VD) Curve (PSNR; IPPP GOP1).
For high spatial complexity and motion activity sequences, variability is significantly lower than the sub-set of sequences with lower spatial and temporal complexity. At the same time, GOP patterns with B frames present higher values of variability regarding GOP patterns without B frames. As frames of type I show lower compression ratio compared to Predicted and Interpolated frames type, the combination of the different types of frames results in the observed higher bit-rate variability.

As the GOP size increases, the amplitude variation regarding the variability increases. This effect is stronger with the video sub-set of lower spatial and temporal complexity sequences. In these cases, motion estimation is very effective resulting in higher compression ratios for P and B pictures comparing to the bits budget of a typical Intra image. B frames, in general, present a small reduction of the variability in sequences with higher complexity. The amplitude of this variation increases while the sequence complexity decreases.

8. Acknowledgment

This work has been supported by “Fundaçao para a Ciência e Tecnologia” and “Programa Operacional Ciência e Inovação 2010” (POCI 2010), co-funded by the Portuguese Government and European Union by FEDER Program.

9. References


www.intechopen.com


IEEE Communications Surveys & Tutorials, Vol. 6, No. 3 (Third Quarter 2004), pp 58-78, ISSN 1553-877X.


This book is intended to attract the attention of practitioners and researchers from industry and academia interested in challenging paradigms of multimedia video coding, with an emphasis on recent technical developments, cross-disciplinary tools and implementations. Given its instructional purpose, the book also overviews recently published video coding standards such as H.264/AVC and SVC from a simulational standpoint. Novel rate control schemes and cross-disciplinary tools for the optimization of diverse aspects related to video coding are also addressed in detail, along with implementation architectures specially tailored for video processing and encoding. The book concludes by exposing new advances in semantic video coding.

In summary: this book serves as a technically sounding start point for early-stage researchers and developers willing to join leading-edge research on video coding, processing and multimedia transmission.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:

© 2011 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License, which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.