RATE AND DISTORTION MODELING ANALYSIS FOR MPEG-4 VIDEO INTRA CODING

Paulo Nunes, Fernando Pereira

Instituto Superior Técnico – Instituto de Telecomunicações
Av. Rovisco Pais, 1049-001 Lisboa, Portugal
Phone: +351.21.8418460; Fax: +351.21.8418472
E-mail: {paulo.nunes, fernando.pereira}@lx.it.pt

ABSTRACT

This paper addresses the problem of rate and distortion modeling in the context of object-based MPEG-4 video encoding by comparing different rate and distortion models for Intra coding in the form of rate-quantization, distortion-quantization and rate-distortion functions. Rate-distortion modeling is an important tool for achieving proper rate-control and thus efficient coding; in this context, the MPEG-4 Visual standard [1] contains an informative annex where a simple rate-quantization quadratic model is proposed. This paper also shows that this model has a higher model fitting error when compared with alternative models with the same number of model parameters.

1. INTRODUCTION

Typically, the aim of a video encoder is to minimize the coded bit rate given a minimum (acceptable) target decoded video quality – *distortion constraint*, or to maximize the perceived video quality given a maximum target bit rate – *rate constraint*. Most of the times, this rate-distortion (RD) tradeoff has to be achieved taking also into account encoder and decoder implementation restrictions – *complexity constraint*, and end-to-end delay restrictions – *delay constraint*.

To achieve an optimal RD trade-off, the rate control mechanisms need to carry on appropriate control actions, e.g., to define the appropriate encoding time instants or to allocate the available bit rate. In this context, two fundamental issues emerge:

- **Rate-distortion modeling** – Targets the design of adequate models for describing the RD behavior associated to the encoding system. These models must capture the statistical characteristics of the source and describe the encoding process as a function of some encoder control parameters, reflecting the lossy encoding RD trade-off.
- **Control process modeling** – Targets the design of a suitable model for describing the encoding process and the corresponding control actions.

This paper addresses the first problem, i.e., the definition of rate and distortion models for object-based video encoding. In this context, the paper analyses a set of rate and distortion models for Intra coding for different rate-control objectives and concludes on their relative merits.

2. RATE AND DISTORTION MODELING

In the context of MPEG-4 video encoding, rate and distortion models characterize the relation between the average number of bits/pixel to code a given Video Object Plane (VOP), the average VOP distortion, and the relevant coding parameters. These models, usually defined in terms of rate-quantization (RQ), distortion-quantization (DQ), and rate-distortion (RD) functions, can play a very important role in real-time video encoding, since they can be used to obtain near optimal operation performance in terms of the RD tradeoff without the drawbacks of having to encode multiple times the same VOP to find the best combination of coding parameters.

For constant bit rate video encoding, the RQ models are useful to compute the quantization parameters to encode each VOP given the bit allocation for the corresponding time instant. Similarly, for approximately constant quality encoding, the DQ models allow to compute the VOP quantization parameters that lead to the target average VOP distortion. Finally, in a multiple video object (VO) encoding scenario, where the rate-control mechanism must keep the quality among the several VOs approximately constant, RD models can be used to guide the bit allocation module in order to produce a bit allocation for the various VOs in the scene that leads to a similar quality.

As pointed out in [2], a fundamental lesson of RD theory is that better performance can be achieved by using a collection of simple models instead of a single all-encompassing model. This principle means that there are typically performance benefits by separating sources of uncertainty and designing a global model as a collection of simple models rather than a single, more complex model. Nevertheless, if the main issue is the modeling error than typically the global model requires several parameters since it is difficult to represent the high-level statistics of the data with a model using a reduced number of parameters. The higher the number of parameters in the model, the higher the number of points needed to estimate the RD model, e.g., the quadratic RQ model proposed in [3] needs at least two points, while the hyperbolic model [4] needs only one point.

Applying to video compression the principle of using a global model that is formed by multiple simpler models requires identifying the main characteristics affecting the RD models. The characteristics/parameters that immediately come out are:

- The coding mode, e.g. Intra, Inter, or Bidirectional.
- The target average compression ratio/average distortion.
- The statistical source data model (type of model), e.g., Laplacian or Gaussian; and model parameters, e.g., mean and variance.

Each coding mode leads typically to coded data with different rate and distortion characteristics, as illustrated in Figure 1 to Figure 3, where the experimental RQ, DQ, and RD functions for one frame of the *Foreman* sequence encoded in Intra and Inter modes are represented (in the case of Inter coding three different reference picture quantization parameter, Q_{ref}, values were used). From these Figures, it is clear that for Intra coding the RQ and DQ functions, exhibit a similar monotonic variation over the whole range of the quantization parameter, i.e., strictly decreasing in the case of the RQ and RD functions and strictly increasing in the case of the DQ function. For Inter coding, however, the RQ, DQ, and RD functions tend to saturate, notably for $Q \gg Q_{ref}$, indicating a clear dependence of the current picture RQ, DQ, and RD functions not only on the quantization parameter of the current picture but also on the quantization parameter of its reference picture. This is shown as different RQ, DQ, and RD curves for different quantization parameter values of the reference picture.

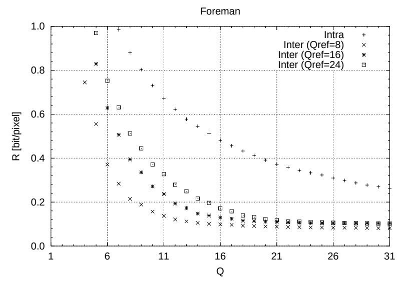

Figure 1 - Experimental RQ function for one frame of the Foreman sequence encoded in Intra and Inter mode

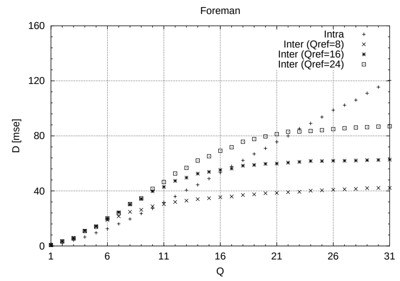

Figure 2 - Experimental DQ function for one frame of the Foreman sequence encoded in Intra and Inter mode

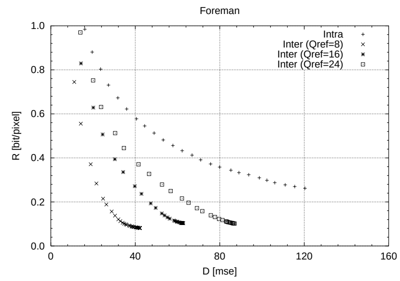

Figure 3 - Experimental RD function for one frame of the Foreman sequence encoded in Intra and Inter mode

It is important to notice that in a RD framework, the number of bits used to encode a given VOP can be divided into two components: one that depends on the quantization parameter(s) used to encode the VOP – *quantizer dependent rate*, and another component that can be considered quantizer independent – *quantizer independent rate*.

In terms of rate control, the *quantizer independent rate* components can be easily estimated from the previous encoding time instants, or, as in the case of the shape and motion bits, they can be obtained before texture encoding during a pre-analysis step, where motion estimation and shape coding are performed [5]. However, in terms of rate and distortion modeling, the *quantizer dependent rate* requires a more careful analysis. This analysis must be done separately for the Intra and Inter coding modes, since, as referred above, these two coding modes exhibit different characteristics in terms of their RD functions. In this paper, this analysis will be restricted to Intra coding. The Inter coding analysis will be developed in the context of future work.

3. RATE AND DISTORTION MODELS ANALYSIS FOR INTRA CODING

In the case of Intra coding, the VOP to be encoded does not depend on other (past or future) VOPs; therefore, its rate and distortion characteristics depend exclusively on the current quantizer parameter(s) and VOP statistics. Since the purpose of this work is to compare several rate and distortion models in order to find for each modeling scenario the model that better approximates the experimental data, it is important to choose analytical models that resemble the typical behavior of the experimental curves as illustrated in Figure 1 to Figure 3.

To evaluate how each model approximates the experimental data, the test sequences *Foreman, Stefan, News, Kayak, Mother and Daughter (M&D), and Football,* in QCIF and CIF formats, at 30 fps, have been encoded with the MPEG-4 Simple Profile without rate control, using only the Intra coding mode for different values of the quantization parameter ($Q \in \{1,\ldots,31\}$).

The model parameters have been estimated for each proposed model and for each encoded picture of each sequence using an iterative non-linear least squares (NLLS) estimate algorithm, more precisely the Levenberg-Marquardt algorithm [6]; moreover, for each model parameter, its minimum, maximum, mean, and standard deviation, measured over all encoded pictures for each sequence, have also been computed.

3.1. Rate-quantization model

The RQ model is used when the primary rate-control objective is to maximize the picture quality given a certain target average bit-rate. The following models have been studied and compared:

RQ model I $$R(Q) = \exp(-a \cdot Q^c + b) \quad (1)$$

RQ model II $$R(Q) = a \cdot \frac{1}{Q^c} + b \quad (2)$$

RQ model III $$R(Q) = \frac{a}{Q^c + b} \quad (3)$$

RQ model IV $$R(Q) = a.\frac{1}{Q^2} + b\frac{1}{Q} + c \quad (4)$$

where *a*, *b*, and *c*, are the model parameters. Table 1 illustrates the RQ model parameter results for the *Foreman* sequence. The less the model parameters depend on the image content, the more robust and useful is the model for rate control purposes. For example, the first three models contain one term where the quantization parameter, *Q*, is raised to a parameter *c*. For model I (1), *c* has an average value of approximately 0.2 for QCIF and

0.1 for CIF; for model II (2) c has an average value of approximately 0.6 for QCIF and 0.8 for CIF; and for model III (3) c has an average value of approximately 1.0 for QCIF and CIF. Since these parameters also exhibit small standard deviations, they can be considered less image dependent than the other model parameters and can be kept constant if a simpler model is aimed. Table 2 presents a new set of model fitting results where some parameters of the models presented above have been kept constant. Notice that by reducing the number of model parameters to two, all models can now be estimated through linear least squares estimates, since they can be rewritten as a straight line equation of the form: $y = a \cdot x + b$.

Notice that the RQ model proposed in the informative Annex L of the MPEG-4 Visual standard [1] corresponds to RQ model IV (4) with $c = 0$ (see Table 2 for the typical model parameters results for the *Foreman* sequence).

3.2. Distortion-quantization model

The DQ model is used when the primary rate-control objective is to minimize the encoded bit-rate given a certain target picture quality. The following models have been studied and compared:

DQ model I $$D(Q) = \exp(a \cdot Q^c + b) \quad (5)$$

DQ model II $$D(Q) = a \cdot (1 - \exp(-b \cdot Q^c)) \quad (6)$$

DQ model III $$D(Q) = a \cdot Q^c + b \quad (7)$$

DQ model IV $$D(Q) = a \cdot Q^2 + b \cdot Q + c \quad (8)$$

where a, b, and c, are the model parameters. Similarly to what occurs for the RQ models, parameter c of DQ models I (5), II (6), and III (7), also exhibits small standard deviations and can be kept constant if a simpler model is aimed. For model I (5), c has an average of value of 0.1; for model II (6) c has an average value of 1.5; and for model III (7) c has an average value of 1.2.

3.3. Rate-distortion model

Although the RQ and DQ models can be used both in single and multiple VO encoding, the RD model is specially useful in the context of multiple VO encoding to guide the bit allocation among the several (arbitrary shape or rectangular) VOs in order to achieve approximately constant quality for all VOs in the scene. The following models have been studied and compared:

RD model I $$R(D) = \exp(-a \cdot (\log_e D - d)^c + b) \quad (9)$$

RD model II $$R(D) = a \cdot \frac{1}{(D-d)^c} + b \quad (10)$$

RD model III $$R(D) = \frac{a}{(D-d)^c + b} \quad (11)$$

RD model IV $$R(D) = a \cdot \frac{1}{D^2} + b \cdot \frac{1}{D} + c \quad (12)$$

where a, b, c, and d, are the model parameters. Notice that RD models I (9), II (10), and III (11), can be obtained analytically combining the appropriate RQ and DQ models. As for the RQ and DQ models, the parameter c for RD models I (9), II (10), and III (11), also exhibits small standard deviations and can be kept constant if a simpler model is aimed. For model I (5), c has an average of value of 1.1; for model II (6) c has an average value of 0.3; and for model III (7) c has an average value of 0.7.

4. MODEL COMPARISON

In order to compare the different models, it is important to choose meaningful comparison criteria. Since the main objective of this model analysis is to obtain a model that closely approximates the experimental data over a wide range of source characteristics, the model comparison criterion used is the minimization of the average model fitting error.

The measure adopted to evaluate how well a given model matches the experimental data for each picture is the standard deviation of the fit (*stdfit*), which is the root mean square of the absolute error between the experimental and estimated data, i.e., for a model with m parameters described by a function, $f(x)$, estimated from N data points, the *stdfit* is defined by

$$stdfit = \sqrt{\sum_{i=1}^{N} (y_i - f(x_i))^2 \Big/ N - m} \quad (13)$$

where $N - m$ is the number of degrees of freedom of the model fitting operation (in the case considered here $N = 31$, corresponding to all possible quantizer values). Thus, the model comparison criterion used is the minimization of the average *stdfit* over all pictures of each sequence. Using the relative error instead of the absolute error does not change the conclusions. More elaborate comparison criteria could be used, notably involving other statistics, such as the maximum *stdfit* over each sequence, and model estimation complexity statistics. Due to space limitations, only the average *stdfit* results were presented.

5. RESULTS

Table 3 to Table 6 present the model fitting results for each studied model and for each test sequence.

Regarding the RQ models with three parameters (see Table 3), models I (1), and II (2), outperform in general models III (3) and IV (4) in terms of the average *stdfit*. For the two parameter models (see Table 4), the average *stdfit* increases between approximately 40% for model IV (4) and 200% for model II (2) for all models; nevertheless, model I (1) still outperforms the other models, notably model IV (4), i.e., the MPEG-4 RQ model [1]. This means that model I (1) would be, in general, the best performing RQ model.

For the DQ models (see Table 5), model II (6) clearly outperforms the other models.

Finally, for the RD models (see Table 6), model II (10) clearly outperforms the other RD models.

6. CONCLUSIONS

This paper studied and compared several rate and distortion models for different rate-control purposes in the form of RQ, DQ, and RD functions capable of accurately modeling the experimental rate and distortion data. In this context, it was also shown that in general the MPEG-4 Visual RQ model [1] has a higher average model fitting error when compared with the alternative models with the same number of parameters, notably RQ model I (1) with two parameters (i.e., $c = c_0$).

REFERENCES

[1] ISO/IEC 14496-2:2001 Coding of Audio-Visual Objects – Part 2: Visual, 2nd Edition, 2001.

[2] M. Effros, “Optimal Modeling for Complex System Design”, *IEEE Sig. Proc.*, vol. 15, no. 6, pp. 51–73, Nov. 1998.

[3] T. Chiang and Y.-Q. Zhang, “A New Rate-Control Scheme using Quadratic Rate-Distortion Model”, *IEEE Trans. on CSVT*, vol. 7, no. 11, pp. 246–250, Feb. 1997.

[4] MPEG Test Model Editing Committee, “MPEG-2 Test Model 5”. ISO/IEC JTC1/SC29/WG11 N400, Apr. 1993.

[5] P. Nunes; F. Pereira; “Scene Level Rate Control Algorithm for MPEG-4 Video Encoding”, *Proc. VCIP 2001*, San Jose, USA, vol. 4310, pp. 194–205, Jan. 2001.

[6] W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing – 2nd Ed.*, Cambridge University Press, 1992.

Table 1 - Foreman (QCIF) RQ model with three parameters

MODEL	PARAM	MIN	MAX	MEAN	STD
I	*a*	2.46	8.00	4.73	1.47
	b	4.18	9.31	6.28	1.38
	c	0.09	0.25	0.16	0.05
II	*a*	3.09	6.28	5.05	0.77
	b	-0.53	-0.09	-0.29	0.13
	c	0.64	0.81	0.69	0.03
III	*a*	4.41	12.45	8.07	2.41
	b	0.33	1.20	0.68	0.29
	c	0.90	1.14	1.00	0.08
IV	*a*	-3.68	-0.96	-2.54	0.64
	b	3.96	9.41	7.24	1.30
	c	-0.07	0.07	0.02	0.05

Table 2 – Foreman (QCIF) RQ model with two parameters

MODEL	PARAM	MIN	MAX	MEAN	STD
I	*a*	3.00	3.57	3.14	0.11
$c = 0.2$	*b*	4.29	5.00	4.68	0.19
II	*a*	3.17	6.34	5.13	0.76
$c = 0.6$	*b*	-0.67	-0.32	-0.49	0.10
III	*a*	3.89	9.42	7.85	1.12
$c = 1.0$	*b*	0.28	0.81	0.66	0.09
II, III	*a*	3.04	5.95	4.88	0.69
$b = 0$	*c*	0.78	0.91	0.81	0.02
IV	*a*	-3.32	-0.82	-2.70	0.47
$c = 0$	*b*	3.84	8.99	7.41	1.07

Table 3 - RQ model with three parameters average stdfit

	SEQ	MODEL			
		I	**II**	**III**	**IV**
QCIF	**Foreman**	0.007	0.018	0.019	0.032
	Stefan	0.030	0.019	0.063	0.119
	News	0.011	0.008	0.022	0.047
	Kayak	0.016	0.024	0.039	0.055
	M&D	0.008	0.018	0.009	0.011
	Football	0.013	0.041	0.025	0.038
AVG QCIF		**0.014**	**0.021**	**0.030**	**0.050**
CIF	**Foreman**	0.011	0.022	0.013	0.013
	Stefan	0.025	0.011	0.039	0.078
	News	0.020	0.013	0.012	0.028
	Kayak	0.018	0.012	0.035	0.052
	M&D	0.011	0.007	0.006	0.008
	Football	0.039	0.020	0.018	0.025
AVG CIF		**0.021**	**0.014**	**0.021**	**0.034**
AVG QCIF + CIF		**0.017**	**0.018**	**0.025**	**0.042**

Table 4 - RQ model with two parameters average stdfit

	SEQ	MODEL				
		I	**II**	**III**	**II,III**	**IV**
QCIF	**Foreman**	0.029	0.047	0.034	0.073	0.046
	Stefan	0.046	0.082	0.062	0.175	0.143
	News	0.034	0.025	0.040	0.057	0.076
	Kayak	0.025	0.030	0.051	0.122	0.061
	M&D	0.019	0.063	0.015	0.042	0.013
	Football	0.026	0.057	0.047	0.119	0.047
AVG QCIF		**0.030**	**0.051**	**0.042**	**0.098**	**0.064**
CIF	**Foreman**	0.034	0.043	0.028	0.057	0.026
	Stefan	0.035	0.114	0.064	0.080	0.124
	News	0.036	0.014	0.052	0.013	0.064
	Kayak	0.043	0.082	0.037	0.086	0.056
	M&D	0.022	0.033	0.012	0.007	0.013
	Football	0.031	0.049	0.028	0.062	0.032
AVG CIF		**0.034**	**0.056**	**0.037**	**0.051**	**0.053**
AVG QCIF + CIF		**0.032**	**0.053**	**0.039**	**0.074**	**0.058**

Table 5 - DQ model average stdfit

	SEQ	MODEL			
		I	**II**	**III**	**IV**
QCIF	**Foreman**	2.593	0.626	1.315	1.675
	Stefan	4.874	1.487	2.476	3.738
	News	1.752	0.766	0.694	1.232
	Kayak	4.052	0.775	2.139	2.608
	M&D	1.489	0.752	0.642	0.699
	Football	2.903	0.948	1.393	1.707
AVG QCIF		**2.944**	**0.892**	**1.443**	**1.943**
CIF	**Foreman**	1.955	0.490	0.974	1.070
	Stefan	2.421	0.521	1.072	1.905
	News	0.834	0.261	0.322	0.666
	Kayak	3.195	0.572	1.598	2.091
	M&D	0.855	0.355	0.322	0.342
	Football	1.958	0.663	0.969	1.125
AVG CIF		**1.870**	**0.477**	**0.876**	**1.200**
AVG QCIF + CIF		**2.407**	**0.685**	**1.160**	**1.572**

Table 6 - RD model average stdfit

	SEQ	MODEL			
		I	**II**	**III**	**IV**
QCIF	**Foreman**	0.034	0.019	0.037	0.189
	Stefan	0.085	0.019	0.078	0.429
	News	0.037	0.016	0.024	0.189
	Kayak	0.062	0.035	0.071	0.273
	M&D	0.022	0.019	0.028	0.098
	Football	0.055	0.036	0.068	0.260
AVG QCIF		**0.049**	**0.024**	**0.051**	**0.240**
CIF	**Foreman**	0.026	0.032	0.035	0.136
	Stefan	0.048	0.022	0.029	0.250
	News	0.021	0.011	0.006	0.082
	Kayak	0.049	0.015	0.052	0.218
	M&D	0.020	0.015	0.011	0.041
	Football	0.038	0.019	0.040	0.152
AVG CIF		**0.034**	**0.019**	**0.029**	**0.147**
AVG QCIF + CIF		**0.041**	**0.022**	**0.040**	**0.193**

ACKNOWLEDGEMENT

We wish to acknowledge the support provided by the European Network of Excellence VISNET (IST Contract 506946).