RATE-DISTORTION ANALYSIS OF NONLINEAR QUANTISERS FOR VIDEO CODERS

Yoo-Sok Saw †, Peter M. Grant‡ and John M. Hannah‡
†Next Generation Telecomm's Div., LGIC R&D Lab.,
533 Hogyea-Dong, Anyang-Si, 430-080, Rep. of Korea
Tel: +82 343 50 2931; Fax +82 343 50 2000
e-mail: yoosoks@lgic.co.kr
‡Dept of Electrical Engineering, University of Edinburgh, Edinburgh, EH9 3JL, UK.
Tel: +44 131 650 5655; Fax: +44 131 650 6554
e-mail: Peter.Grant@ee.ed.ac.uk

ABSTRACT

Quantisation is used in video coders such as MPEG in association with rate control scheme to regulate the data rate of compressed video bit stream entering the transmission buffer. When the transmission data rate is limited the quantiser has a crucial effect on video data rate and video quality. The quantiser step size is generally determined by a linear non-adaptive method with respect to the buffer occupancy. In this paper, we investigate in the framework of rate-distortion theory two adaptive nonlinear quantiser control functions, sigmoidal and unimodal, which achieve superior video rate control performance while maintaining similar video quality to the linear case. Their performance for video rate fluctuation has also been analysed in both analytic and experimental ways.

1 INTRODUCTION

The quantiser controls the data rate (or buffer occupancy) of the current coded picture in a feedback loop, typically in a buffer-based MPEG (Motion Picture Experts Group) video encoder. This buffer-based scheme can be used both for constant and variable bit rate video applications. Much research for constant bit rate (CBR) applications has focused on two different approaches; adaptive quantisation [1, 2, 3] and control theoretic approaches [4] which enable the video encoder to control its output data rate to match the fixed rate channel.

Several different quantiser control functions have been proposed so far. They can be classified into linear [5], piecewise linear [2, 4] and nonlinear [1, 3]. Recently, two representative nonlinear control functions have been reported; sigmoidal and unimodal which will be analysed in the following sections.

For block transform coding operating on 1st-order stationary Gauss-Markov sources, it is difficult to solve accurately the relationship between the entropy of transformed picture data and the quantisation distortion, since the entropy and the distortion are related to each other through the quantiser (step size). In feedback buffer-quantiser rate control schemes the occupancy is the only variable parameter to control or adjust the quantiser step size in the rate-distortion sense. However, the rate-distortion analysis derives the resulting distortion for a given data rate not from the quantiser step size or the buffer occupancy but from the variance of the data to be quantised. The distortion in practice needs to be determined in terms of the quantiser step size derived as a function of the occupancy. This paper attempts to bridge this gap and improve the design of quantiser control functions.

Although digital video rarely has Gaussian statistics, the rate-distortion function, $D(R)$, is interpreted as the upper bound to the actual rate-distortion performance, since, it is known that the rate-distortion function for non-Gaussian sources, $D(R)_N$, has the following relationship with that for Gaussian sources, $D(R)_G$, that is, $D(R)_N \leq D(R)_G$ [6]. With this practical use of the rate-distortion function, we analyse the two quantiser control functions for $D(R)_G$ rather than $D(R)_N$ for mathematical tractability. In this paper $D(R)_G$ will be expressed in terms of the buffer occupancy not the data rate. Furthermore, the functions will be evaluated with respect to the control performance over the occupancy fluctuation. Simulation results will also be presented, to indicate which quantiser control function is more suitable for realistic video data.

2 Nonlinear quantisation and its rate-distortion theoretic background

Although the distribution of DCT video data may differ considerably from Gaussian pdf, we use the following rate-distortion function [6];

$$D(R)_G = \gamma_x^2 2^{-2R} \sigma_x^2 \qquad (1)$$

where R is the available bit rate. $D(R)_G$ is the rate-distortion function for a memoryless zero-mean Gaussian source $N(0, \sigma_x^2)$ with the variance σ_x^2. γ_x^2 is the spectral flatness measure which represents the correlation in the source data. For simplicity, it is set to 1, i.e. implying a memoryless source since it is assumed to be constant. In the following sections, two performance criteria will be examined; the rate-distortion functions for the sigmoidal and the unimodal functions and the bit rate fluctuation property of these two functions.

3 Rate-Distortion functions, $D_s(R)$ and $D_u(R)$

3.1 Preposition 1: the buffer occupancy

Let r be the buffer occupancy, with a range of real values 0 to 1, be represented as a function, $r = f(B, R_T, \sigma_x^2)$. Here, B and R_T, respectively, represent the buffer size in bits and a given data rate in bits per pixel which are set to constant values in most CBR applications. Therefore, it can be simplified to $r = f(\sigma_x^2)$, which means that the buffer occupancy r is the function of the source variance, σ_x^2. We introduce the short-term source variance, $\sigma_{x\Delta t}^2$, which represents the variance during the short period of time, Δt, in relation to the long-term variance, σ_x^2. For a specific R_T value, Eqn. 1 reduces to $D(R)_G = 2^{-2R_T}\sigma_x^2 = H\sigma_x^2$ where H is a constant. This signifies that the distortion is given by the only variable σ_x^2, i.e. $D = g(\sigma_x^2)$ where $g(\ )$ is an arbitrary function. Let $r(\Delta t)$ be the short-term occupancy. Then, we can describe the occupancy, r, with respect to the variance σ_x^2; if $\sigma_{x\Delta t}^2 > \sigma_x^2$ then $r(\Delta t)$ increases, if $\sigma_{x\Delta t}^2 < \sigma_x^2$ then it decreases, otherwise, it remains unchanged. Thus, the occupancy equation can be given as follows:

$$r = r_0 \left(\frac{\sigma_{x\Delta t}^2}{\sigma_x^2} \right)^b \tag{2}$$

where r_0 is the initial occupancy at $t = t_0$. The index b represents the behaviour of the buffer occupancy, which is determined by the buffer size. If the buffer size is small the b value should be large enough to correctly describe the variation of r. Note that the initial occupancy r_0 is multiplied by the variance ratio rather than added to it. Since $\sigma_{x\Delta t}^2 = k\sigma_x^2$, the real value k should be positive. If k is close to 1, Eqn. 2 results in $r = r_0$ for an arbitrary value of b.

3.2 Preposition 2: the relationship between the bit rate and the occupancy

Let R and q_* be the available bit rate depending on the occupancy r and the quantisation step size for the picture $*$, respectively. For arbitrary functions, $f(\ )$ and $g(\ )$, the relationship between the bit rate and the quantisation step size can be given as follows:

$$R = f(q_*) = f(g(r)) = R_T \frac{1}{q_*^\alpha}, \quad (\alpha \geq 1) \tag{3}$$

where α is a positive integer to determine the nonlinearity of Eqn. 3. Note that R is different from R_T; R_T is a constant channel rate, however, R is the available bit rate determined by the quantiser step size, Eqn. 3. If Eqn. 3 has a quadratic term, it will be given as follows:

$$R = R_T \left(\frac{u}{q_*} + \frac{v}{q_*^2} \right) \tag{4}$$

where u and v are the parameters which must be optimised. Bigger α values introduce more nonlinear relationship between R and q_*. However, since the resulting rate-distortion function is given by the similar variable r, α is set to 1 for simplicity. The function, $g(r)$, specifies the relationship between the occupancy and the quantiser step size. This relationship can be either linear or nonlinear and we discuss two nonlinear functions: sigmoidal and unimodal.

3.3 Analysis of $D_s(R)$ and $D_u(R)$

From Eqn. 3 the quantiser step size, q_*, is replaced with q_s or q_u according to the definitions of sigmoidal and unimodal [3] functions as follows:

$$q_s = \frac{1}{1 + \exp(-\beta(r_s - 0.5))} \tag{5}$$

$$q_u = r_u^{1/\beta} \tag{6}$$

where r_s and r_u are corresponding occupancy variables to the sigmoidal and unimodal functions, respectively. β determines the curvature of the control functions. Their graphical representations are shown in Fig. 1 and 2, respectively. The curvature parameter, β, is substituted for a transmission parameter of rate balance, R_b, which represents the ratio of the number of bits transmitted for a short period of time to the allocated number of bits.

$$R_b = \frac{coded\_bits|_{\Delta t}[bits] \ \times \ picture\_rate[frame/s]}{channel\_rate[bit/s] \times \ number\_of\_pictures|_{\Delta t}[frames]} \tag{7}$$

A constant scaling factor c is multiplied by R_b, Eqn. 5, i.e. cR_b is used instead of R_b, in order to make the shape of the sigmoidal function appropriate for quantiser control. Here, c is set to 10 so that the curvature of the sigmoidal functions appears to be a skewed-"S" shape. In Figs. 1 and 2 the curves are shown for representative R_b values, i.e. 0.5, 1, 1.5 and 3. The upper bound of R_b depends on the buffer size.

From Eqns. 1, 3, 5 and 6, the rate-distortion functions, $D_s(R)$ and $D_u(R)$, for the sigmoidal and unimodal functions can be derived and from Eqn. 3, $R = \frac{R_T}{q_*}$, replacing β with cR_b in Eqns. 5 and 6:

$$q_s = \frac{1}{1 + \exp(-cR_b(r_s - 0.5))} \tag{8}$$

$$q_u = r_u^{1/R_b} \tag{9}$$

Since this derivation procedure aims to represent the rate-distortion functions in terms of the occupancy and the rate balance, in Eqn. 2 the long-term variance, σ_x^2, and the initial occupancy, r_0, are set to 1 for simplicity and eventually we achieve the following rate-distortion functions expressed in terms of $r_{s,u}$ and R_b:

$$D_s(r_s, R_b) = 2^{-2R_T[1+\exp(-cR_b(r_s-0.5))]} r^{1/b} \tag{10}$$

$$D_u(r_u, R_b) = 2^{-2\left(\frac{R_T}{r_u^{1/R_b}}\right)} r^{1/b} \tag{11}$$

b is a constant and r_s and r_u represent the sigmoidal and unimodal occupancies, respectively. Fig. 3 shows, in 3 dimensions, the resulting rate-distortion functions depending on r and R_b. The z-axis represents the relative distortion with respect to σ_x^2. Fig. 3(a) and (b) refer to the transmission with a low bit rate, 768 kbit/s or 0.202 bits per pixel (bpp).

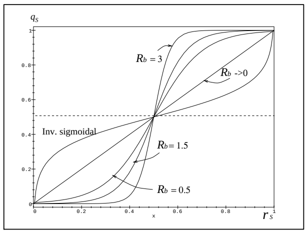


Figure 1: The sigmoidal quantiser control function.

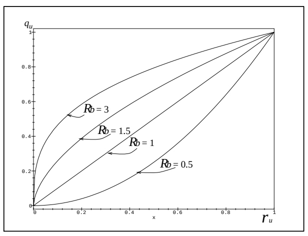
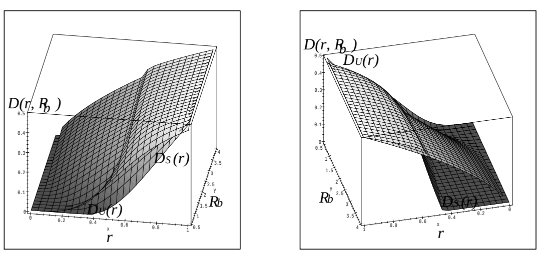


Figure 2: The unimodal quantiser control function.



(a) 0.202 bpp (front) (b) 0.202 bpp (rear)

Figure 3: 3-dimensional representation of the rate-distortion function, $D(r, R_b)$, at 768 kbit/s. Surface plots variation with rate balance, R_b, and occupancy, r.

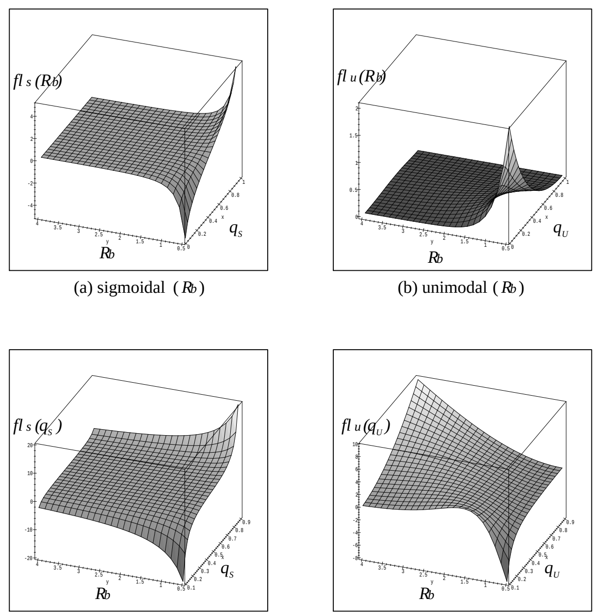
The area of interest in these plots is the region with lower R_b, say, smaller than 2 and the occupancy, r, less than 80%, Fig. 3(a) (front view). In this area $D_u(r_u, R_b)$ appears to lie below $D_s(r_s, R_b)$. This suggests that the unimodal function can provide more reduced distortion than the sigmoidal function. On the other hand, if R_b is higher than 2, the sigmoidal function shows a better distortion profile, Fig. 3(b) (rear view), where the occupancy is lower than 50%. However, this case is not considered as normal coding operation. The results appear identical for the two different transmission rates, 768 and 2048 kbit/s.

4 Controllability of the occupancy fluctuation

Fluctuation in the buffer occupancy is the second important performance criterion, since it has direct influence on the performance of rate control. The occupancy fluctuation can be defined as the instantaneous variation of the occupancy, i.e. the second derivative of its transmission parameters; the rate balance and quantiser step size. The first derivative of the occupancy is the average occupancy for a relatively large interval of the parameter. Since the quantiser step size was expressed as a function of $r_{s,u}$ and R_b, Eqn. 5 and 6, the fluctuation, fl_s for the sigmoidal and fl_u for the unimodal, become the second partial derivatives as follows:

$$\frac{\partial^2 r_s}{\partial R_b^2} = fl_s(R_b) \quad , \quad \frac{\partial^2 r_u}{\partial R_b^2} = fl_u(R_b)$$
$$\frac{\partial^2 r_s}{\partial q_s^2} = fl_s(q_s) \quad , \quad \frac{\partial^2 r_u}{\partial q_u^2} = fl_u(q_u) \qquad (12)$$

These fluctuation equations can also be depicted with 3-dimensional plots, as shown in Fig. 4.



(a) sigmoidal (R_b) (b) unimodal (R_b)

(c) sigmoidal (q_s) (d) unimodal (q_u)

Figure 4: 3-dimensional representation of the rate balance fluctuation, $fl(R_b)$, and the quantiser step size variation, fl_q.

The z-axis represents the relative fluctuation in terms either of the rate balance or of the quantiser step size. The R_b value ranges from 0.5 to 4. The z-axis values are theoretical ones which are useful to assess the performance of these functions. For the R_b fluctuations, $fl_s(R_b)$ and $fl_u(R_b)$, the unimodal function exhibits much smaller variation than the sigmoidal functions for the same ranges of the rate balance and the quantisation scale. Note that the z-axis ranges of $fl_s(R_b)$ and $fl_s(q_s)$ (sigmoidal) are much wider than those of the unimodal. For the quantisation scales, q_s and q_u, $fl_u(q_u)$ appears to have a steady and smaller fluctuation profile in comparison to $fl_s(q_s)$. Thus, this leads to a judgement that the unimodal control function should offer superior control of the video rate fluctuation.

5 Numerical results

To verify these analysis results, the quantisation parameter equation of the MPEG2 TM5 was replaced by the two quantisation control equations, i.e. Eqn. 5 and 6. The rate balance, R_b, Eqn. 7, is determined by:

$$R_b = \frac{c\_bits*}{t\_bits*} \tag{13}$$

where c_bits* and t_bits* represent the numbers of coded and target bits allocated for the previous picture $*$, respectively. The value of R_b was clipped at 2. Two different channel rates were tested: 768 and 1024 kbit/s. Two video sequences were used: "Starwars" (300 frames) and "JFK" (330 frames). The image format is MPEG SIF in which the picture size is 352 pixels (row) by 240 lines (column). Table 1 summarises the performance, showing mean and standard deviation (std. dev.) for the measures: peak signal-to-noise ratio (PSNR) for the reconstructed luminance signals and the occupancy for the 2 frames buffer.

Video sequence	Channel rate [kbits/s]	Rate control	PSNR [dB]		Occupancy [%]	
			mean	std. dev.	mean (peak)	std. dev.
Starwars	768	TM5	31.67	2.62	59 (119)×	18.6
		Sigm.	31.64	2.60	54 (96)	10.2
		Unim.	31.65	2.61	40 (65)	7.5
	1024	TM5	33.27	2.69	41 (82)	10.8
		Sigm.	33.03	2.68	46 (69)	4.7
		Unim.	33.04	2.69	33 (54)	5.4
JFK	768	TM5	34.10	4.17	61 (273)×	44.7
		Sigm.	34.10	4.08	79 (356)×	77.4
		Unim.	34.04	4.27	39 (123)×	17.5
	1024	TM5	35.70	4.13	39 (145)×	23.0
		Sigm.	35.65	4.11	46 (133)×	14.0
		Unim.	35.66	4.13	30 (68)	10.8

Table 1: Mean and standard deviation of the peak SNR and occupancy performance measures for 2 video sequences at 3 transmission rates with 2 rate control techniques where × highlights buffer overflow.

TM5 represents the MPEG2 TM5 evaluation model. Sigm. and Unim. represent the modified TM5 versions based on the sigmoidal and unimodal functions, discussed previously. Peak occupancy greater than 100% means that buffer overflow has occurred. The Unim. exhibits virtually the same PSNR as the other two methods for all the channel rates. On the other hand, the occupancy shows very different values. TM5 is least able to control the occupancy fluctuation. Unim. outperforms Sigm. for all the channel rates by showing far smaller fluctuation in occupancy with the same video quality. Note that Sigm. shows far larger fluctuations than Unim. and TM5 at 768 kbit/s. This implies that Sigm. does not adequately control dramatic rate fluctuation when the rate balance has large values due to low channel rate and dramatic scene changes. This agrees with the analysis, presented in Section 4.

These simulations were conducted to investigate the performance of the nonlinear quantiser control functions by simply replacing the TM5's quantisation equation with the sigmoidal and unimodal functions. However, they can perform better when an integral rate control scheme is employed. The feed-forward, nonlinear network-based or fuzzy logic-based scheme is an attractive control scheme [7].

6 Conclusion

In this paper two nonlinear quantiser control functions for MPEG video encoders have been analysed and compared in a rate-distortion theoretic way. A unimodal function was shown to provide better performance than a sigmoidal function in both the rate-distortion aspect and the controllability over video rate fluctuation. This was also confirmed by conducting simulations on an MPEG2 TM5 encoder for realistic video sequences. This provides a useful approach to the analytic evaluation of buffer-quantiser control algorithms.

Acknowledgements

Yoo-Sok Saw acknowledges the support of *The British Council* and *LG Information and Communications Ltd., Korea*.

References

[1] J. Zdepski and D. Raychaudhuri, "Statistically based buffer control policies for constant rate transmission of compressed digital video," *IEEE Trans. Commun.*, vol. 39, pp. 947 – 957, 1991.

[2] N. Farvardin and J. W. Modestine, "Adaptive buffer-instrumented entropy-coded quantiser performance for memoryless sources," *IEEE Trans. Information Theory*, vol. IT-32, no. 1, pp. 9–22, 1986.

[3] Y. Saw, P. M. Grant, and J. M. Hannah, "Feed-forward buffering and rate control based on scene change parameters for MPEG video coder," in *EUSIPCO'96*, vol. II, pp. 727–730, EURASIP, 10 - 13, September, 1996.

[4] J.-P. Leduc, "Bit-rate control for digital TV and HDTV codecs," *EURASIP Signal Processing: Image Communication*, vol. 6, pp. 25–45, 1994.

[5] L.-W. Lee, J.-F. Wang, J.-Y. Lee, and C. C. Chen, "On the error distribution and scene change for the bit rate control of MPEG," *IEEE Trans. Consumer Electronics*, vol. 39, no. 3, pp. 545–554, 1993.

[6] N. Jayant and P. Noll, *Digital coding of Waveforms: Principles and Applications to Speech and Video.* Prentice-Hall, 1984.

[7] Y. Saw, P. M. Grant, J. M. Hannah, and B. Mulgrew, "Nonlinear predictive rate control for constant bit rate mpeg video coders," in *ICASSP'97*, vol. 4, pp. 2641–2644, IEEE, 21-24, April, 1997.