COEFFICIENT THRESHOLDING AND OPTIMIZED SELECTION OF THE LAGRANGIAN MULTIPLIER FOR NON-REFERENCE FRAMES IN H.264 VIDEO CODING

Pontus Carlsson[1], F. Pan[2], Liang-Tien Chia[3]

[1]Popwire Technology
[2]Institute for Infocomm Research
[3]Nanyang Technological University

pontus@popwire.com, efpan@i2r.a-star.edu.sg, asltchia@ntu.edu.sg

ABSTRACT

A new strategy to select the Lagrangian multiplier for non-reference frames is proposed, with the objective to maximize the average PSNR. The selection is based on the observation that non-reference frames should be optimized using a Lagrangian multiplier equivalent to the negative slope of the global R-D curve. This curve is not known *a priori* and a simple approximation is presented. Further, a new criterion for transform coefficient thresholding is proposed, based on the current frame type and Lagrangian optimization.

The new scheme is shown to improve the PSNR between 0.35 and 1.12 dB compared to the H.264 Test Model. The average improvement for all sequences is 0.63 dB, or equivalently[1] a bitrate reduction of 11%.

1. INTRODUCTION

Rate-distortion optimization for video compression has proven to yield substantial improvements in perceived quality and the H.264 Test Model is designed around this concept. In this Test Model, R-D optimization is performed for each frame individually and the problem of distributing the bits among frames to achieve the global optimum is not addressed. This is usually the case in video coding as the strong dependence between frames makes global optimization very complex. In this paper we address the problem of how to select the local coding parameters for non-reference frames in order to maximize the global performance.

The process of dropping transform coefficients that have a non-zero value after quantization is usually referred to as *coefficient thresholding*. By doing this in a rate-distortion framework, substantial gain in PSNR can be obtained for JPEG and previous MPEG video standards [2, 3]. In these standards, static variable length coding of transform coefficients is combined with run-length coding of zeros. This makes it possible to, for each coefficient, determine its rate and distortion, and algorithms based on dynamic programming can be designed to achieve R-D optimum. However, the entropy coding scheme in H.264 is more complex. It utilizes binary arithmetic coding and context adaptation to dynamically update the probability distributions. Performing R-D optimized thresholding for this scheme is difficult. We target the simpler problem of whether to discard or to keep the coefficients in an 8×8 block consisting of four transform blocks. This decision is integrated into the R-D optimization process of the H.264 Test Model and a block is discarded only if this results in a lower R-D cost.

[1]Based on interpolation of the R-D curves as described in [1]

2. LAGRANGIAN MULTIPLIER FOR NON-REFERENCE FRAMES

The objective of video coding is to minimize the distortion of the coded sequence at a given target bitrate. This can be formulated as

$$\min D(R) \qquad \text{subject to} \qquad R \leq R_c \tag{1}$$

where D, R and R_c denotes distortion, bitrate and target bitrate respectively. The solution to this minimization problem can be obtained using Lagrangian optimization, where the cost function J is defined as

$$J(R) = D + \lambda R \tag{2}$$

Minimizing this function results in an optimal solution to (1) for a particular value of R_c [4, 5]. The minimum value of $J(R)$ can be found by setting its derivative to zero which yield

$$\lambda = -\frac{\partial D(R)}{\partial R} \tag{3}$$

Thus, at optimality, λ corresponds to the negative slope of the R-D curve.

In general, rate and distortion can be assumed additive [6] and (2) can be formulated as

$$J(R_0, \ldots, R_{N-1}) = \sum_{i=0}^{N-1} D_i(R_0, \ldots, R_{N-1}) + \lambda \sum_{i=0}^{N-1} R_i \tag{4}$$

where i is the frame number, R_i is the number of bits for frame i and N the total number of frames in the sequence. This equation can be minimized by setting all its partial derivatives with respect to R_i to zero. For any frame k we have

$$\frac{\partial J}{\partial R_k} = \frac{\sum_{i=0}^{N-1} \partial D_i}{\partial R_k} + \lambda \frac{\sum_{i=0}^{N-1} \partial R_i}{\partial R_k} = \frac{\sum_{i=0}^{N-1} \partial D_i}{\partial R_k} + \lambda \tag{5}$$

and we obtain the equation system

$$\begin{cases} \frac{\partial J}{\partial R_0} = \frac{\sum_{i=0}^{N-1} \partial D_i}{\partial R_0} + \lambda = 0 \\ \vdots \\ \frac{\partial J}{\partial R_k} = \frac{\sum_{i=0}^{N-1} \partial D_i}{\partial R_k} + \lambda = 0 \\ \vdots \\ \frac{\partial J}{\partial R_{N-1}} = \frac{\sum_{i=0}^{N-1} \partial D_i}{\partial R_{N-1}} + \lambda = 0 \end{cases} \quad (6)$$

As

$$\frac{\sum_{i=0}^{N-1} \partial D_i}{\partial R_k} + \lambda = \frac{\partial D_k}{\partial R_k} + \frac{\sum_{i \neq k} \partial D_i}{\partial R_k} + \lambda = 0 \quad (7)$$

for each k, we conclude that

$$\frac{\partial D_k}{\partial R_k} = -\lambda - \frac{\sum_{i \neq k} \partial D_i}{\partial R_k} \quad (8)$$

should hold at optimality.

In practical video coding it is impossible to consider overall global dependencies and the optimization must be performed more or less locally. A frame level Lagrangian cost function can be defined as

$$J_k = D_k + \lambda_k R_k \quad (9)$$

which is minimized analogously to (2) yielding

$$\lambda_k = -\frac{\partial D_k}{\partial R_k} \quad (10)$$

However, our objective is to minimize the global R-D cost and a question that arises is

How should the local Lagrangian multiplier λ_k be selected in order to minimize the global R-D cost?

Noting that (8) and (3) should hold at global optimum, (10) can be modified as

$$\lambda_k = -\frac{\partial D_k}{\partial R_k} = \lambda + \frac{\sum_{i \neq k} \partial D_i}{\partial R_k} = -\frac{\partial D}{\partial R} + \overbrace{\frac{\sum_{i \neq k} \partial D_i}{\partial R_k}}^{I(k)} \quad (11)$$

The optimal local Lagrangian multiplier thus reflects the impact of the local rate on the global distortion which is represented by $I(k)$.

In inter frame video coding there is a strong dependence between frames and $I(k)$ will in general have a large negative value. This is due to the fact that higher quality reference frames also improve the quality of subsequently coded frames. In general $I(k)$ is difficult to approximate. However, a special case is when there is no dependence between frames. $I(k)$ is equal to zero. Lagrangian optimization under similar independence assumptions is the most commonly studied in literature [4, 6].

We observe that this special case is valid for non-reference frames as their rate does not affect the distortion of other frames[2] and conclude that *for non-reference frames, the local Lagrangian multiplier should be set to the negative slope of the global R-D curve.*

[2]When using rate control this is not entirely true. Here we consider strict Lagrangian optimization where the rate cannot be explicitly controlled but is merely a result of the selected Lagrangian multiplier.

2.1. Lagrangian Multiplier Selection in H.264 Test Model

In [5], a strong connection between the local Lagrangian multiplier and the quantization parameter QP was found experimentally as

$$\lambda_k \approx 0.85 \times 2^{(QP-12)/3} \quad (12)$$

In the H.264 Test Model the Lagrangian multiplier for I and P-frames ($\lambda_{I,P}$) is set according to this relation. For non-reference frames, or equivalently B-frames in the H.264 Test Model[3], it is modified as

$$\lambda_B = 4\lambda_{I,P} \quad (13)$$

where $\lambda_{I,P}$ is derived from a given QP according to (12).

2.2. Proposed Lagrangian Multiplier Selection

For non-reference frames we suggest the local Lagrangian multiplier to be equal to the global Lagrangian multiplier. We then select QP using (12), which has been shown to be a near optimal mapping between the quantization parameter and the Lagrangian multiplier [5]. By solving QP from this equation we obtain

$$QP_B = 3\log_2(\frac{\lambda}{0.85}) + 12 \quad (14)$$

where QP_B denotes the quantization parameter for B-frames. Thus, in contrast to the H.264 Test Model, QP for B-frames is set to be adaptive to the negative slope of the global R-D curve.

3. APPROXIMATION OF THE GLOBAL RATE-DISTORTION FUNCTION

In general, the global R-D curve is not known *a priori* and is subject to approximation. Experimentally, the following equation has been found to well approximate the global PSNR curve for a variety of sequences and bitrates (Fig.1)

$$PSNR \approx C_1 + 10\sqrt{2}\log(R) \quad (15)$$

where C_1 is a constant dependent on the source statistics. Using the definition of PSNR, this can be shown equivalent to the distortion D being approximated as

$$D \approx \frac{C}{R^{\sqrt{2}}} \quad (16)$$

where C is another constant dependent on source statistics. As λ should be equal to the negative slope of the R-D curve we have

$$\lambda \approx -\frac{\partial}{\partial R}\left(\frac{C}{R^{\sqrt{2}}}\right) = \frac{C\sqrt{2}}{R^{1+\sqrt{2}}} \approx \frac{D\sqrt{2}}{R} \quad (17)$$

where (16) is used to obtain the final approximation. Thus, λ can be approximated using the total rate and distortion. Unfortunately these quantities are also not known *a priori*. Assuming that the intra period is known and the GOP structure is fixed, we use

$$\begin{cases} R_{avg} \approx \frac{\bar{R_I} + \bar{R_{P,B}}(N_{GOP}-1)}{N_{GOP}} \\ D_{avg} \approx \frac{1}{k}\sum_{i=0}^{k-1} D_i \end{cases} \quad (18)$$

[3]Unlike previous MPEG standards, B-frames in the H.264 standard are not restricted to being non-reference frames. However, in the H.264 Test Model they are.

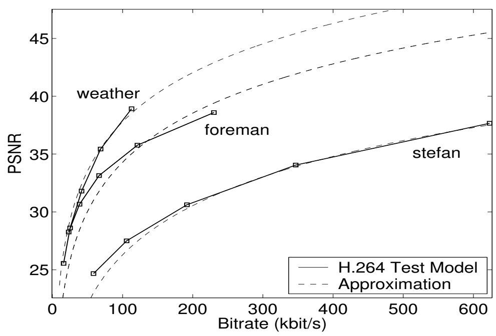

Fig. 1. Approximation of the global rate-distortion function. The operational rate-distortion functions (solid) have been obtained by encoding the sequences using fixed QP.

to estimate the average rate and distortion, denoted as R_{avg} and D_{avg}. N_{GOP} is the total number of frames in a GOP. $\bar{R}_I$ and $\bar{R}_{P,B}$ denote the current average rate for intra and inter frames respectively and k is the current frame number.

4. MODIFICATION OF THE COEFFICIENT THRESHOLDING CRITERION

4.1. Thresholding in H.264 Test Model

In the H.264 Test Model, all transform coefficients within an 8×8 block will be set to zero if their total cost is below a threshold which is set to 4. Noting that the transform blocksize in H.264 is 4×4, this thresholding can be described as

$$\sum_{i=0}^{3}\sum_{j=0}^{15} c(i,j) \leq 4 \tag{19}$$

where i is the transform block index in raster scan order and j is the transform coefficient index in zig-zag scan order. $c(i,j)$ is a cost function defined in the H.264 Test Model as

$$c(i,j) = \begin{cases} \infty & \text{if coeff}(i,j) > 1 \\ 3 & \text{if } j = 0 \text{ and coeff}(i,j) = 1 \\ 2 & \text{if } 1 \leq j \leq 2 \text{ and coeff}(i,j) = 1 \\ 1 & \text{if } 3 \leq j \leq 5 \text{ and coeff}(i,j) = 1 \\ 0 & \text{otherwise} \end{cases}$$

where coeff(i,j) is the transform coefficient value after quantization. Thus, only coefficients equal to 1 will be discarded. The coefficient index, or frequency, determines its cost. For example if all non-zero coefficients within an 8×8 block have the value one and their associated coefficient indices are greater than 5, the total cost will be zero and the block will be discarded. If two out of the four DC coefficients equals to 1, the cost will be 6 and the block is retained.

4.2. R-D Optimized Thresholding

We note that

- P-frames are used as reference frames and global performance might be improved by using a lower threshold.

Table 1. Test Conditions

MV Resolution	1/4 pel
Hadamard	ON
RDO	ON
Search Range	±32
Reference Frames	1
Symbol Mode	CABAC
GOP Structure	IBBP
Intra Period	0

Table 2. Rate and PSNR comparison between the H.264 Test Model with fixed QP and the proposed scheme.

	QP = 28		QP = 36	
Sequence	PSNR	Rate(%)	PSNR	Rate(%)
akiyo.qcif	+0.68	-2.60	+0.44	+0.13
foreman.qcif	+0.49	-0.78	+0.26	-0.13
mad.qcif	+0.60	-3.13	+0.31	+2.29
mobile.qcif	+0.79	-0.20	+0.25	-0.86
news.qcif	+0.96	-0.21	+0.86	+2.48
salesman.qcif	+1.25	0.49	+0.41	+0.53
silent.qcif	+1.01	-2.29	+0.63	-1.12
stefan.qcif	+0.49	-0.14	+0.63	0.94
weather.qcif	+1.40	-0.09	+0.69	-2.98

- Thresholding can be rate-distortion optimized.

Based on these observations we evaluate the R-D cost twice instead of once for each 8×8 block. The rate and distortion is first obtained for the case of keeping all coefficients and then for the case of discarding all coefficients. The one that has the lowest R-D cost is then selected for final coding. For non-reference frames the R-D cost is evaluated using the same Lagrangian multiplier as for mode selection, i.e. the approximated negative slope of the global R-D curve obtained from (17). For reference frames a lower value is used, as they are more important and their coding quality affects many other frames. In the experiments, we scale the original Lagrangian multiplier with 0.2. This has the effect that coefficients are only discarded if their rate is very high compared to their energy.

5. RESULTS

The new scheme has been integrated into version 6.1c of the H.264 Test Model together with the H014 rate control algorithm described in [7]. For the experiments we use frame level rate control, i.e. QP is the same for all macroblocks within a frame. The setup of the encoder is given in Table 1. In Table 2 the H.264 Test Model using fixed QP is compared to the proposed scheme. We note a significant increase in PSNR for similar bitrate. Further, the improvement tend to be higher for lower QP, a trend also observed in Figure 2 - 4.

In Table 3 the average improvement of PSNR and bitrate is calculated according to [1]. This calculation is based on interpolation of the R-D curve and, opposed to Table 2, the two columns should be regarded as equivalent. There is either the increase in PSNR or the decrease in rate - not both at the same time.

For more experimental results we refer to [8].

Table 3. Average improvement of PSNR and rate compared to fixed QP calculated according to [1]. The columns should be regarded as equivalent in the sense that there is either the increase in PSNR or the decrease in rate.

Sequence	PSNR	Rate(%)
akiyo.qcif	+0.47	-7.53
foreman.qcif	+0.35	-7.31
mad.qcif	+0.36	-7.82
mobile.qcif	+0.49	-10.54
news.qcif	+0.88	-14.66
salesman.qcif	+0.58	-11.42
silent.qcif	+0.82	-15.82
stefan.qcif	+0.61	-11.15
weather.qcif	+1.12	-14.40
Average	+0.63	-11.18

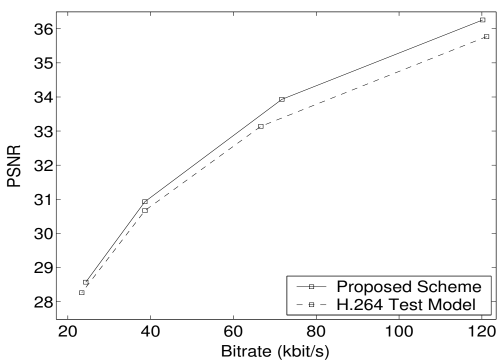

Fig. 2. Foreman (QCIF). Average PSNR improvement +0.35 dB.

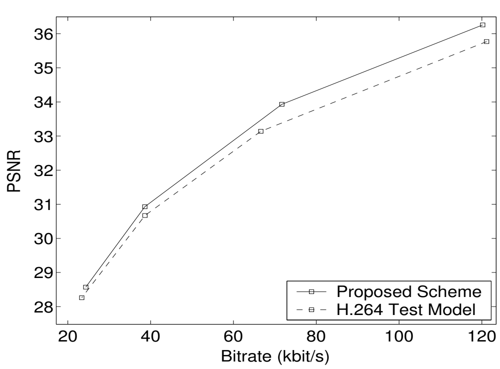

Fig. 3. Mobile (QCIF). Average PSNR improvement +0.49 dB.

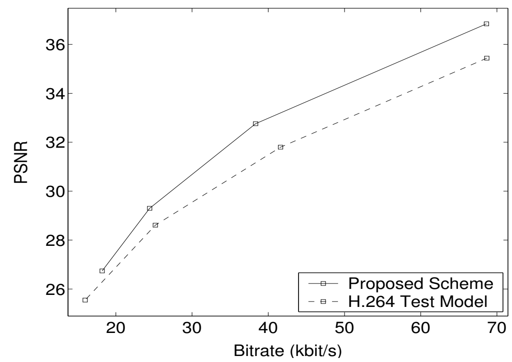

Fig. 4. Weather (QCIF). Average PSNR improvement +1.12 dB.

6. CONCLUSIONS

A method for selecting the Lagrangian multiplier and associated quantization parameter for non-reference frames in H.264 video coding has been proposed. Together with rate-distortion optimized thresholding the new scheme increases the PSNR significantly for all sequences tested. The average improvement for all sequences is 0.63 dB, or equivalently a bitrate reduction of 11%.

7. REFERENCES

[1] Gisle Bjontegaard, "Calculation of average PSNR differences between RD-curves," in *VCEG-M33*, Apr. 2001.

[2] Kannan Ramchandran and Martin Vetterli, "Rate-distortion optimal fast thresholding with complete JPEG/MPEG decoder compatibility," in *IEEE Transactions on Image Processing*, Sept. 1994, vol. 3, pp. 790–794.

[3] Matthew Crouse and Kannan Ramchandran, "Joint thresholding and quantizer selection for transform image coding: entropy-constrained analysis and applications to baseline JPEG," in *IEEE Transactions on Image Processing*, Feb. 1997, vol. 6, pp. 285–297.

[4] Yair Shoham and Allen Gersho, "Efficient bit allocation for an arbitrary set of quantizers," in *IEEE Transactions on Acoustics, Speech and Signal Processing*, Sept. 1988, vol. 36, pp. 1445–1453.

[5] Gary J. Sullivan and Thomas Wiegand, "Rate-distortion optimization for video compression," in *IEEE Signal Processing Magazine*, Nov. 1998, vol. 15, pp. 74–90.

[6] Thomas Wiegand, Heiko Schwarz, Anthony Joch, Faouzi Kossentini, and Gary J. Sullivan, "Rate-constrained coder control and comparison of video coding standards," in *IEEE Transactions on Circuits and Systems for Video Technology*, July 2003, vol. 13, pp. 688–703.

[7] Z. G. Li, Wen Gao, F. Pan, S. Ma, K. P. Lim, G. Feng, X. Lin, S. Rahardja, H. Lu, and Y. Lu, "Adaptive rate control with HRD consideration," in *ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6 (JVT-H014)*, May 2003.

[8] Pontus Carlsson, "Transform coefficient thresholding and Lagrangian optimization for H.264 video coding," M.S. thesis, Linköpings Universitet, Feb. 2004.