ADVANCED LAGRANGE MULTIPLIER SELECTION FOR HYBRID VIDEO CODING

Xiang Li†, Norbert Oertel†, Andreas Hutter† and André Kaup**

*Chair of Multimedia Communications and Signal Processing, University of Erlangen-Nuremberg
†Networks & Multimedia Communications, Siemens Corporate Technology

ABSTRACT

The Lagrangian multiplier based rate-distortion optimization has been proved to be an effective way in hybrid video coding. In this paper, an advanced Lagrange multiplier selection method is presented. Based on Laplace distribution, the variance of transformed residuals is introduced into the rate and distortion models. Moreover, inspired by the ρ-domain method, the percentage of non-zeros among quantized residuals is also considered to refine the rate model. Thanks to the more accurate models, the proposed method is able to adaptively optimize the encoding process for videos with different properties. It outperforms the algorithm used in the reference software of H.264/AVC. According to the simulation, a gain up to 1.1 dB was achieved.

1. INTRODUCTION

In hybrid video codec design, rate-distortion optimization is a critical part that greatly affects the overall performance. The target of such optimization is to minimize the distortion D for a given rate R_c by appropriate selections of coding parameters, namely

$$\begin{gathered}\min\{D\}\\ \text{subject to } R \leq R_c\end{gathered} \tag{1}$$

However, it is hard to solve such a constrained problem directly. Therefore, (1) is usually converted to (2) by the Lagrange multiplier method.

$$\begin{gathered}\min\{J\}\\ \text{where } J = D + \lambda \cdot R\end{gathered} \tag{2}$$

Where J is the Lagrangian cost function and λ is the so-called Lagrange multiplier. Consequently, how to determine λ becomes a key problem in rate-distortion optimization.

In order to achieve a good λ, researchers developed many methods. Among them, the algorithm proposed in [1, 2] is widely used. Supposing R and D to be differentiable everywhere, the minimum of the Lagrangian cost J is given by setting its derivative to zero, i.e.

$$\frac{\mathrm{d}J}{\mathrm{d}R} = \frac{\mathrm{d}D}{\mathrm{d}R} + \lambda = 0 \tag{3}$$

yielding

$$\lambda = -\frac{\mathrm{d}D}{\mathrm{d}R} \tag{4}$$

In fact, (4) indicates that λ corresponds to the negative slope of the rate-distortion curve, that is λ can be perfectly determined by the models of R and D. Assuming a sufficiently high rate environment, the R model can be derived as shown in (5) according to the typical high-rate approximation curve for entropy-constrained scalar quantization [3].

$$R(D) = a \log_2\left(\frac{b}{D}\right) \tag{5}$$

Where a and b are two constants. For the D model, if the same "high rate" assumption is maintained, the source probability distribution can be approximated as uniform within each quantization interval [3], leading to

$$D = \frac{(2Q)^2}{12} = \frac{Q^2}{3} \tag{6}$$

Note that here the quantizer value Q is half the distance of the quantizer reproduction levels. Putting (5) and (6) into (4), the final λ can be determined by

$$\lambda = -\frac{\mathrm{d}D}{\mathrm{d}R} = c \cdot Q^2 \tag{7}$$

where c is a constant which is experimentally suggested to be 0.85 [2]. In summary, this method, referenced as High Rate λ selection (HR-λ) for convenience, is practical and efficient. Therefore it was adopted into the reference softwares not only by H.263 [4][5], but also by H.264/AVC [6][7], the state-of-art video coding standard. However, there are some drawbacks with this algorithm. First, λ is only related to Q and no property of the input signal is considered, which means that it can not adapt to different videos dynamically. More important, the "high rate" assumption is not realistic all the time, which will result in a poor performance for both rate and distortion models, especially in the case of low bit-rate coding.

In order to achieve adaptivity, Chen and Garbacea proposed an adaptive λ estimation method based on the ρ-domain method [8]. Defining ρ as the percentage of zeros among quantized residuals, the R and D can be rewritten as (8) and (9), respectively [9].

$$R(\rho) = \theta \cdot (1 - \rho) \tag{8}$$

$$D(\rho) = \sigma^2 \cdot e^{-\alpha(1-\rho)} \tag{9}$$

Where θ and α are coding constants, σ^2 is the variance of transformed residuals. Considering (4) at the same time, Chen finally derived λ as shown in (10) [8].

$$\lambda = \beta \cdot (\ln(\sigma^2/D) + \delta) \cdot D/R \tag{10}$$

Where β and δ are both coding constants. In summary, such method represents a first shot at adaptivity. Thanks to the introduction of the variance of transformed residuals, it is able to adapt videos dynamically. However, it decouples quantization information Q from λ, which is sometimes inconvenient, such as in the case of rate control. More important, the rate model is a kind of approximation [10], which may greatly degrade the performance of the algorithm because of the error propagation by the λ's direct derivation from R and D.

To improve the accuracy of the models while achieving the adaptivity at the same time, an advanced Lagrange multiplier selection method is presented in this paper. Instead of "high rate" assumption, the new rate and distortion models for the proposed algorithm are based on the model of Laplace distribution of transformed residuals. Moreover, inspired by the rate expression of the ρ-domain method, the percentage of non-zeros among the quantized residuals is also considered to refine the rate model. Taking advantage of the new models, the proposed method achieves a better performance than HR-λ while keeping the direct relationship with Q. The rest of this paper is organized as follows. First, the proposed method is detailedly depicted in Section 2. Then, simulation results and related discussions are presented in Section 3. Finally, the conclusion and future work are given in Section 4.

2. LAGRANGE MULTIPLIER SELECTION

In this section, the rate and distortion models for the proposed method are first derived from Laplace distribution. Then the Lagrange multiplier selection method is discussed. Finally, the implementation details are presented.

2.1. Rate and Distortion Models

It has been well accepted that, after motion estimation (or intra prediction for intra-frame coding), the transformed residuals follow a Laplace distribution f_L [11]. Normally, it can be supposed as a zero-mean distribution, i.e.

$$\begin{aligned} f_L(x) &= \frac{\lambda_L}{2} e^{-\lambda_L |x|} \\ \lambda_L &= \frac{\sqrt{2}}{\sigma} \end{aligned} \tag{11}$$

where x represents the transformed residual and σ is their standard deviation which indicates the property of the input video. Because of the one-to-one mapping between σ and λ_L, the latter will be used instead in the following expressions for simplicity. And it should be also noted that λ_L is a distribution parameter which is different from the Lagrange multiplier λ mentioned in Section 1.

Based on Laplace distribution, the entropy of the quantized transformed residuals H can be modeled as (12) for uniform reconstruction quantizers, such as that in H.264/AVC.

$$\begin{aligned} H &= -P_0 \cdot \log_2 P_0 - 2\sum_{n=1}^{\infty} P_n \cdot \log_2 P_n \\ P_0 &= \int_{-(Q-F)}^{Q-F} f_L(x)dx \\ P_n &= \int_{nQ-F}^{(n+1)Q-F} f_L(x)dx \end{aligned} \tag{12}$$

Where Q is the quantization interval and F is the rounding offset. Similar with H, distortion D can be written as

$$D = \int_{-(Q-F)}^{Q-F} x^2 f_L(x)dx + 2\sum_{n=1}^{\infty} \int_{nQ-F}^{(n+1)Q-F} (x-nQ)^2 f_L(x)dx \tag{13}$$

In theory, the entropy expression (12) can be used to represent the rate directly. Nevertheless, the accuracy can be further improved. Assuming that pixels are coded independently, (12) describes the average entropy for each pixel. However, in video coding standards, such as H.264/AVC [6], a kind of run-length coding technique is applied to the quantized block level, which means successive zeros in zigzag order will be coded together to save bits. Consequently, if the successive zeros take a big percentage, such as in low bit-rate inter-frame coding, (12) will not be sufficiently accurate. Inspired by ρ-domain method, the percentage of non-zeros among quantized residuals is included into the rate model to solve this problem. Normally in video coding, the average entropy for zeros is much smaller than that for non-zeros. Therefore, only the entropy for non-zeros are counted in the rate model, i.e.

$$R = S \cdot (1 - P_0) \cdot H \tag{14}$$

where P_0 is the probability of quantized zeros defined in (12), and S is a constant to compensate the errors resulted from both non-ideal Laplace distribution and the neglect on entropy for zeros.

2.2. Advanced Lagrange Multiplier Selection

In this sub-section, the proposed advanced Lagrange Multiplier selection method is derived. And then a mathematical discussion is made on the method.

First, R can be easily obtained by plugging (11) and (12) into (14).

$$\begin{aligned} R = &\frac{S \cdot e^{-\lambda_L(Q-F)}}{\ln 2}(-(1-e^{-\lambda_L(Q-F)})\ln(1-e^{-\lambda_L(Q-F)}) \\ &+ e^{-\lambda_L(Q-F)}(\ln 2 - \ln(1-e^{-\lambda_L Q}) - \lambda_L F + \frac{\lambda_L Q}{1-e^{-\lambda_L Q}})) \end{aligned} \tag{15}$$

Similarly, D can be achieved by putting (11) into (13).

$$D = \frac{e^{\lambda_L F}(2\lambda_L Q + \lambda_L^2 Q^2 - 2\lambda_L^2 QF) + 2 - 2e^{\lambda_L Q}}{\lambda_L^2(1-e^{\lambda_L Q})} \tag{16}$$

Intuitively, λ_L is an inherent property of the input video. For the case of no shot transition, it is reasonable to be regarded as a constant during a short time, such as several frames. Thus

(4) turns to

$$\lambda = -\frac{\mathrm{d}D}{\mathrm{d}R} = -\frac{\partial D/\partial Q}{\partial R/\partial Q} \tag{17}$$

If (15) and (16) are put into (17), the proposed λ can be determined. Unfortunately, such complete expression is too long to be presented here. Nevertheless, in practice, it can be easily computed by mathematical softwares or numerical methods. Of course, a lookup table can also be used to save computations.

To achieve some simplicity, series expansion can be applied to (17). For example, in H.264/AVC inter-frame coding where the rounding offset F equals to $Q/6$ [7], (17) becomes

$$\lambda = \frac{\ln 2}{S}\left(\frac{7}{18}Q^2 - \frac{8 + 35(\ln\frac{10}{3} - \ln Q - \ln\lambda_L)}{108}Q^3\lambda_L\right) + \mathrm{O}(\lambda_L^2) \tag{18}$$

Consequently, when λ_L is close to zero and Q is small, the high order terms of λ_L in (18) are neglectable. Moreover, when λ_L approaches zero, the Laplace distribution will reduce to the uniform distribution, which yields the limit of (18)

$$\lim_{\lambda_L \to 0} \lambda = \frac{7\ln 2}{18S} \cdot Q^2 \tag{19}$$

In mathematics, (19) exactly matches (7). Therefore, it can be concluded that HR-λ is actually a special case of the proposed method.

2.3. Implementation Details

The proposed Lagrange multiplier selection method, or in short Laplace λ selection (Lap-λ), is a general method for the rate-distortion in hybrid video coding. For example, it was embedded into the Main Profile of H.264/AVC environment. There are several practical aspects related to the implementation worth mentioning. Most of them are on the derivations of parameters for Lap-λ calculation.

First, Q in (15) and (16) is determined by the quantization parameter qp in H.264/AVC.

$$Q = 2^{\frac{(qp-12)}{6}} \tag{20}$$

And to be in line with the reference software [7], the rounding offset F in (15) and (16) are set to $Q/3$ and $Q/6$ for intra-frame and inter-frame coding, respectively.

Second, the factor S in (15) can be calculated by comparing the limit of Lap-λ (19) to HR-λ (7). Keeping c as 0.85 as in (7), S for inter-frame coding in H.264/AVC will be

$$S = \frac{7\ln 2}{18c} = 0.317 \tag{21}$$

Similarly, the S for intra-frame coding can be derived as 0.181. Of course, S can also be determined by experiments.

Third, λ_L in (15) and (16) is calculated according to the variance of transformed residuals of each frame. Since it is supposed to be constant over a short time, the λ_L of the previous frame is used as the estimation to the current frame. For the case of H.264/AVC Main Profile, as 4:2:0 sampling is used, only the variance of luminance component is counted in the λ_L calculation because of the dominant role of luminance signal. In fact, the proposed method can be easily extended to other samplings, such as 4:4:4, by counting the related chrominance signal into λ_L.

Finally, the proposed λ can be calculated according to (15) (16) and (17). In practice, to avoid being trapped into ill solutions, the range of the final λ by (17) should be limited. In the current implementation, it is set to between 20% and 500% of the corresponding value by HR-λ.

It should be also noted that, for the very first intra-frame and inter-frame, as no λ_L information available, HR-λ is used instead of Lap-λ.

Table 1. Simulation Results

sequences	qp	HR-λ PSNR (dB)	HR-λ Rate (kbit/s)	HR-λ Aver λ	Lap-λ PSNR (dB)	Lap-λ Rate (kbit/s)	Lap-λ Aver λ	$\triangle$PSNR (dB)
container (qcif, 30 Hz)	28	37.64	27.55	34.3	37.47	23.76	42.72	0.66
	32	35.36	13.80	86.4	35.06	10.89	193.49	
	36	33.02	7.99	217.6	32.68	6.10	842.21	
	40	31.02	5.31	548.3	30.57	4.00	2577.88	
news (qcif, 30 Hz)	28	37.85	64.51	34.3	37.67	60.86	41.84	0.29
	32	35.24	37.73	86.4	34.94	34.06	121.01	
	36	32.72	22.08	217.6	32.25	18.38	358.47	
	40	30.49	13.14	548.3	29.83	9.57	1497.2	
foreman (qcif, 30 Hz)	28	37.13	96.95	34.3	36.93	91.25	59.82	0.11
	32	34.94	52.05	86.4	34.54	46.22	194.39	
	36	32.84	31.27	217.6	32.16	25.82	677.19	
	40	30.85	20.01	548.3	29.88	15.29	2200.77	
paris (cif, 30 Hz)	28	36.59	504.28	34.3	36.53	497.40	38.14	0.09
	32	33.95	280.96	86.4	33.72	263.33	110.29	
	36	31.40	152.77	217.6	31.02	134.37	346.93	
	40	29.17	84.78	548.3	28.55	67.05	1124.56	
silent (qcif, 30 Hz)	28	37.03	69.94	34.3	36.92	67.09	44.7	0.28
	32	34.60	41.34	86.4	34.38	37.46	127.44	
	36	32.38	24.22	217.6	31.99	19.58	399.65	
	40	30.45	13.67	548.3	29.86	9.63	1359.2	
mobile (cif, 30 Hz)	28	34.90	1492.04	34.3	34.97	1523.45	40.06	0.02
	32	32.05	711.35	86.4	31.98	694.46	121.15	
	36	29.40	322.32	217.6	29.11	291.75	401.25	
	40	27.06	166.27	548.3	26.44	142.38	1306.72	
tempete (cif, 30 Hz)	28	35.92	952.59	34.3	35.95	962.35	39.29	0.04
	32	33.30	439.80	86.4	33.15	416.54	119.68	
	36	30.84	199.13	217.6	30.49	174.62	387.73	
	40	28.67	99.97	548.3	28.12	81.57	1493.83	
Average	-	-	-	-	-	-	-	0.21

3. SIMULATIONS AND DISCUSSIONS

The proposed method was tested with JM 11.0 [7], the most recent reference software of H.264/AVC. The simulation environment strictly conformed to the suggestion on coding efficiency test [12, 13] by JVT (Joint Video Team by ISO/IEC MPEG & ITU-T VCEG) except the frame structure, i.e. one I frame followed by 99 P frames with no frame skipping. Since the I frame in the simulation is always coded with HR-λ, only P frame information was collected in the result.

Table 1 shows the simulation result. Since for H.264/AVC Main Profile, the rate-distortion process affects not only the compression on luminance but also that on chrominance, the

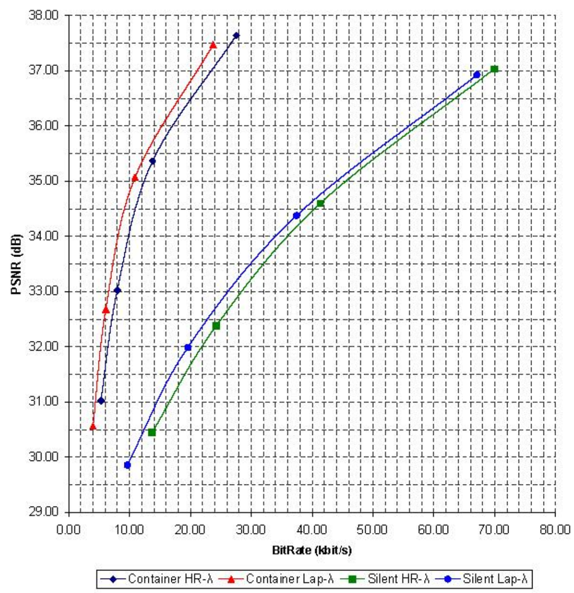

Fig. 1. rate-distortion curves for container and silent.

PSNR in the table is calculated according to

$$P = \frac{1}{6}(4P_Y + P_U + P_V) \tag{22}$$

where P_Y, P_U and P_V are PSNRs for luminance and two chrominance components, respectively. For $\triangle$PSNR, it is obtained with the PSNR calculator recommended by JVT [14]. In fact, it indicates the average PSNR difference between two RD-curves. Thus it can be used to as an objective evaluation on the whole performance.

Thanks to the more accurate models, the proposed method Lap-λ shows a superior performance over HR-λ. The average gain for the seven test sequences is 0.21 dB, and up to 0.66 dB gain is achieved for the sequence "container". As mentioned above, such gain obtained by the calculator for each sequence is a kind of average. The peak gain is actually even bigger. For example, Fig. 1 shows that the peak gains for the "container" and "silent" are up to 1.1 dB and 0.5 dB, respectively.

However, for videos with big movements, rotations or zooming, such as the "mobile" and "tempete", the gain by Lap-λ is very small. Such result can be explained by the models. Normally, the residual variance for the fast videos is much bigger than that for slow ones because of less precise prediction. Therefore the distribution in such case is close to the assumption of uniform distribution, which results that HR-λ can also work fine. Thus HR-λ and Lap-λ share a similar performance. Contrariwise, for slow videos, the assumption of HR-λ gets failed and Lap-λ provides a better result.

4. CONCLUSION AND FUTURE WORK

In this paper, a Laplace distribution based Lagrange multiplier selection method Lap-λ for rate-distortion optimization in hybrid video coding is presented. Considering the variance of transformed residuals in both rate and distortion models, the proposed algorithm achieves not only the accuracy but also the adaptivity. According to the simulation, Lap-λ outperforms the current technique HR-λ in the reference software of H.264/AVC and a gain up to 1.1 dB was observed. More important, mathematical analysis shows that HR-λ is actually a special case of the proposed Lap-λ.

On the other hand, there still exists rooms for improvement. First, the current rate model is directly derived from quantized residuals and no side information, such as motion vector, is considered. In theory, a further gain should be achievable if such side information is taken into account. Second, although the proposed method works with B-frame as well, an even better performance is expected if a good joint optimization among P and B frames is available.

5. ACKNOWLEDGMENT

The authors would like to thank Mr. Roy Ghandour for his helpful work on the paper.

6. REFERENCES

[1] G.J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE Signal Process. Mag.*, vol. 15, no. 6, pp. 74–90, 1998.

[2] T. Wiegand and B. Girod, "Lagrange multiplier selection in hybrid video coder control," in *IEEE Int. Conf. on Image Processing*, 2001, pp. 542–545, vol.3.

[3] H. Gish and J. Pierce, "Asymptotically efficient quantizing," *IEEE Trans. Inf. Theory*, vol. 14, no. 5, pp. 676–683, 1968.

[4] ITU-T, *Video Coding for Low Bitrate Communication*, ITU-T Rec. H.263 Version3 (H.263++). 2000.

[5] ITU-T/SG16/Q15-D-65, "Video Codec Test Model, Near Term, Version 10 (TMN-10), Draft 1," April 1998.

[6] JVT, *Advanced Video Coding (AVC) - 3rd Edition*, ITU-T Rec. H.264 and ISO/IEC 14496-10 (MPEG-4 Part 10). 2004.

[7] JVT, "H.264/AVC reference software (JM11.0)," http://iphome.hhi.de/suehring/tml/, August 2006.

[8] L. Chen and I. Garbacea, "Adaptive lambda estimation in lagrangian rate-distortion optimization for video coding," in *Visual Comm. and Image Processing, Jan 17-19 2006*, 2006.

[9] Z.H. He and S.K. Mitra, "Optimum bit allocation and accurate rate control for video coding via ρ-domain source modeling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 10, pp. 840–849, 2002.

[10] Z.H. He and S.K. Mitra, "A linear source model and a unified rate control algorithm for DCT video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 11, pp. 970–982, 2002.

[11] E.Y. Lam and J.W. Goodman, "A mathematical analysis of the DCT coefficient distributions for images," *IEEE Trans. Image Process.*, vol. 9, no. 10, pp. 1661–1666, 2000.

[12] G. Sullivan, "Recommended simulation common conditions for h.261coding efficiency experiments on low-resolution progressive-scan source material," in *ITU-T VCEG SG16 Q.6 14th Meeting, VCEG-N81*, Santa Barbara, USA, Sept. 2001.

[13] A. Hallapuro, M. Karczewicz, and H. Malvar, "Low complexity transform and quantization - Part II: Extensions," in *Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, JVT-B039*, 2002.

[14] G. Bjontegaard, "Calculation of average psnr differences between rd-curves," in *ITU-T VCEG SG16 Q.6 13th Meeting, VCEG-M33*, Austin, Texas, USA, Apr. 2001.