A NOVEL TWO-PASS VBR CODING ALGORITHM FOR THE H.264/AVC VIDEO CODER BASED ON A NEW ANALYTICAL R-D MODEL

Jianfei Huang[1], Jun Sun[2], and Wen Gao[2]

[1]*Graduate University of Chinese Academy of Science, Beijing, 100080, China*
[2]*Institute of Digital Media, School of Electronic Engineering and Computer Science, Peking University, Beijing 100871, China*

ABSTRACT

In this paper, the generalized Gaussian distribution is employed first to model the 16(4*4) integer transform coefficients of the residue image in H.264 videos. Then the distortion-rate function of generalized Gaussian model is analyzed and an effective and flexible rate-distortion (R-D) model is developed to approximate the actual R-D function. Finally, an effective variable bit rate (VBR) algorithm for H.264/AVC is proposed, which adopts two-pass encoding to achieve a constant video quality. Experimental results show that under the same average bit rate, the proposed algorithm achieves 94% to 99% reduction in PSNR variation compared to the model of JM9.8.

Index Terms— Two-pass coding, VBR, rate-distortion, statistical model, bit allocation, H.264

1. INTRODUCTION

Over the past few decades, digital video compression technique was extensively used in video communication, TV broadcast, video on demand, and digital storage. Several international industry standards have been established, such as MPEG-1/2/4, H.263 and H.264. The newest H.264/AVC video coding standard developed by JVT doubled the coding efficiency in comparison with earlier standards [1]. It is prospected that the H.264/AVC videos would be prevalent for its expressive coding efficiency.

For all video encoders, rate control plays an important role. The output bit stream can be either constant bit rate (CBR) or variable bit rate (VBR). In many bandwidth constrained applications, CBR encoding is widely adopted because of its low complexity and easy implementation. However, CBR encoding is difficult to achieve the consistent visual quality, and usually has low coding efficiency. Compared to CBR, VBR encoding is able to optimize the bit allocation, and provide constant quality.

Commonly, VBR encoding needs two-pass encoding. Characteristics of the entire video sequence are collected and analyzed in the first pass, then the sequence could be optimally re-encoded in the second pass [2]-[4]. However, previous works might be imperfect for the lack of a precise rate-distortion model.

In this paper, we present a two-pass VBR coding algorithm for H.264/AVC, using our new R-D model. The model is developed based on the statistical analysis of the integer transform coefficients with generalized Gaussian distributions (GGD) [5].

The remainder of this paper is organized as follows: In Section 2, we employ GGD to model integer transform coefficients and develop a new rate-distortion model. Our two-pass VBR coding algorithm is proposed in Section 3. Section 4 presents the experimental results. Finally, the conclusions are drawn in Section 5.

2. ANALYTICAL RATE-DISTORTION MODEL

2.1. Statistical Analysis of DCT Coefficients with GGD

Generalized Gaussian distribution [5] is a nice model for the transform coefficients. The probability density function (PDF) of zero-mean GGDs can be described as follows:

$$p(x) = \frac{\alpha\eta(\alpha,\beta)}{2\Gamma(1/\alpha)}\exp\{-[\eta(\alpha,\beta)\,|\,x\,|]^{\alpha}\} \quad (1a)$$

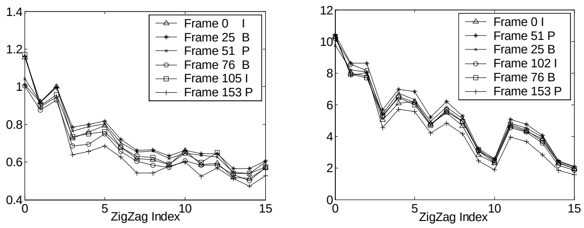
with

$$\eta(\alpha,\beta) = \beta^{-1}\left[\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)}\right]^{1/2} \quad (1b)$$

where $\alpha > 0$ is the shape parameter describing the exponential rate of decay, β is a positive quantity representing a scale parameter.

Fig. 1 shows the estimations of α and β for all the 4*4 integer transform coefficients respectively in the I, P and B pictures in the sequence "Foreman" with CIF. The root variance β is normalized with the transform scaling factor [6]. It can be seen from Fig. 1 that not only the root variance decreases typically, but also the shape parameter shows the same trend of decrease in the Zig-Zag order. And

This work was supported by National Key Technology R&D Program under contract No. 2006BAH02A10 and China Postdoctoral Science Foundation (No. 20060400030).



(a) Shape Parameter α (b) Normalized Root variance β

Fig. 1. GGDs' parameter estimation of 16 integer transform coefficients in the frame 0, 25, 51, 76, 105, 153 of the "Foreman" Sequence

most integer transform coefficients have the shape value in the range 0.5-1.0.

2.2. Quantization scheme of H.264

In H.264/AVC, the main principle of the quantization scheme can be expressed as:

$$Z = \text{int}\left(\frac{|W| + f \bullet \Delta}{\Delta}\right) \cdot \text{sign}(W) \quad (2a)$$

and

$$W' = \Delta \cdot Z \,, \quad (2b)$$

where equation (2a) describes the quantization, equation (2b) describes the inverse quantization. In equation (2a), the input signal *W* is mapped to a quantization level *Z* , *Δ* is the quantization step size, *f* is the rounding control parameter, the function *int()* rounds float data to the nearest integer towards minus infinity. The mapping of the quantization level *Z* to the reconstructed signal *W'* is described in the inverse quantization step of equation (2b).

The quantization and inverse quantization can also be represented by equation (2a) and (2b). In the reference model of H.264/AVC, the rounding control parameter is *f=1/3* for Intra mode and *f=1/6* for Inter mode. Fig. 2 shows the two quantization scheme.

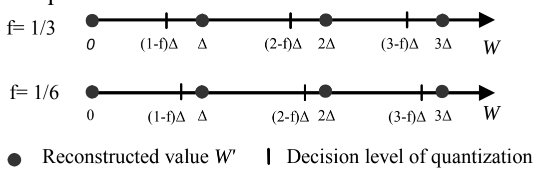


Fig. 2 The quantization and inverse quantization in H.264/AVC coding with two different parameters *f*.

2.3. Rate-Distortion Analysis of GGD under Quantization Scheme of H.264

It is important to analyze the R-D function of generalized Gaussian distributions under the quantization scheme of H.264 [6]. For simplicity, we rewrite the PDF of GGD (equation (1)) as following:

$$p(x) = \tfrac{k_1(\alpha)}{\beta} \exp(-[k_2(\alpha)\tfrac{|x|}{\beta}]^\alpha) \quad (3)$$

where $k_1(\alpha) = \alpha[\Gamma(3/\alpha)]^{1/2} / 2[\Gamma(1/\alpha)]^{3/2}$ and $k_2(\alpha) = \left[\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)}\right]^{1/2}$. For an arbitrary variable $\delta > 0$, let the quantization step size $\Delta_\beta = \beta\delta$, the dead-zone control parameter is *f*. If the probability at the reconstruction level $k\Delta_\beta$ is $P_k(\Delta_\beta)$, then we obtain

$$P_k(\Delta_\beta) = \begin{cases} \int_{-(1-f)\beta\delta}^{(1-f)\beta\delta} \frac{k_1(\alpha)}{\beta} \exp(-[k_2(\alpha)\frac{|x|}{\beta}]^\alpha) dx \\ \quad = \int_{-(1-f)\delta}^{(1-f)\delta} k_1(\alpha) \exp(-[k_2(\alpha)\,|\,x\,|]^\alpha) dx, \ \text{k=0} \\ \int_{(k-f)\beta\delta}^{(k+1-f)\beta\delta} \frac{k_1(\alpha)}{\beta} \exp(-[k_2(\alpha)\frac{|x|}{\beta}]^\alpha) dx \\ \quad = \int_{(k-f)\delta}^{(k+1-f)\delta} k_1(\alpha) \exp(-[k_2(\alpha)\,|\,x\,|]^\alpha) dx, \ \text{k} \ge 1 \\ P_{-k}(\Delta_\beta), \qquad \text{k} \le -1 \end{cases} . \quad (4)$$

For a specific shape parameter α and constant *f*, it's easy to see that $P_k(\Delta_\beta)$ is just determined by the variable δ . The entropy rate of the quantization can be calculated by

$$H(\Delta_\beta) = \sum_{k=-\infty}^{\infty} -P_k(\Delta_\beta) \log_2 P_k(\Delta_\beta) \,. \quad (5)$$

So the entropy rate is also just determined by δ . That is $H(\Delta_\beta) = f_\alpha(\delta)$, which is a decreasing function of δ .

The quantization scheme is symmetric about zero, so the MSE of the quantization can be computed as

$$D(\Delta_\beta) = 2\sum_{k=0}^{+\infty} D_k(\Delta_\beta) \quad (6)$$

where

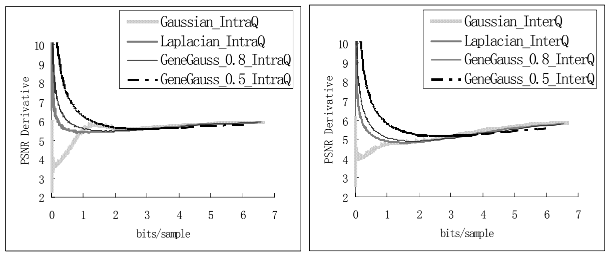
$$D_k(\Delta_\beta) = \begin{cases} \int_0^{(1-f)\beta\delta} (x - k\beta\delta)^2 \frac{k_1(\alpha)}{\beta} \exp(-[k_2(\alpha)\frac{|x|}{\beta}]^\alpha) dx \\ \quad = \beta^2 \int_0^{(1-f)\delta} (x - k\delta)^2 k_1(\alpha) \exp(-[k_2(\alpha)\,|\,x\,|]^\alpha) dx, \ \text{k=0} \\ \int_{(k-f)\beta\delta}^{(k+1-f)\beta\delta} (x - k\beta\delta)^2 \frac{k_1(\alpha)}{\beta} \exp(-[k_2(\alpha)\frac{|x|}{\beta}]^\alpha) dx \\ \quad = \beta^2 \int_{(k-f)\delta}^{(k+1-f)\delta} (x - k\delta)^2 k_1(\alpha) \exp(-[k_2(\alpha)\,|\,x\,|]^\alpha) dx, \ \text{k} \ge 1 \end{cases} \quad (7)$$

That is, $D(\Delta_\beta) = \beta^2 g_\alpha(\delta) = \beta^2 g_\alpha(f_\alpha^{-1}(H(\Delta_\beta)))$. Then if using PSNR as the distortion criteria, we can obtain

$$\begin{aligned} PSNR(\Delta_\beta) &= 20\log_{10} 255 - 10\log_{10} D(\Delta_\beta) \\ &= 20\log_{10}(255/\beta) - 10\log_{10}(g_\alpha(f_\alpha^{-1}(H(\Delta_\beta)))) \end{aligned} . \quad (8)$$

From (8) it's easy to see that $d_{PSNR(\Delta_\beta)} / d_{H(\Delta_\beta)}$ is not relative to the root variance β .

We have done some computer simulations of (8) with different shape parameter α . Fig. 3 shows the derivative comparison of the distortion-rate (D-R) functions with the



(a) Intra mode (b) Inter mode

Fig. 3. Derivative comparison of the distortion-rate functions with the GGD shape parameter 0.5, 0.8, 1.0 (Laplacian) and 2.0 (Gaussian).

GGD shape parameter 0.5, 0.8, 1.0 (Laplacian) and 2.0 (Gaussian). Fig. 3(a) shows the Intra quantization scheme (*f=1/3*). Fig. 3(b) shows the Inter quantization scheme (*f=1/6*). Generally, the derivative of the D-R functions decreases to the traditional number 6.02 dB/bit. For some large shape parameters such as 2.0, the derivative of the D-R functions increases as the bit rate increases in the actual coding rate range (Below 1.0 bit/sample).

2.4. Rate-Distortion Model of H.264

Considering the distribution of actual integer transform coefficients in Fig. 1(a), we could assume that the derivative of distortion-rate function (PSNR criterion) decreases continuously, as the rate increases. With above analysis, we get a heuristic R-D model as follows [7]:

$$PSNR(R) = aR + A - (A-B)/(1+bR) \tag{9}$$

where R is the bit rate per sample, B is the PSNR value when R equals to 0, a and A is the asymptote parameter, b is the parameter controlling the approach of the actual rate-distortion function (RDF) to the asymptote. The parameters a and b can also be set as the constants for a coarse model.

Besides the *PSNR-R* model, we design a simple *PSNR-Q* model as well. It is known that, at high bit rates, the mean-square error (MSE) versus quantization factor (Q) relation of an entropy-coded uniform quantizer can be approximated by [8]:

$$MSE \approx Q^2 / c\,.$$

While

$$PSNR = 10\log_{10}(255^2 / MSE)$$

the *PSNR-Q* model can be concluded as

$$PSNR(Q) = c\log_{10} Q + d \tag{10}$$

where c, d are parameters, Q is the quantization factor Qstep.

3. PROPOSED TWO-PASS VBR CODING ALGORITHM

The proposed algorithm mainly consists of three parts: the first-pass encoding, an offline processing, and the second-pass encoding.

The purpose of the first-pass encoding is to get features such as bit numbers, Qp and PSNR values of the entire video sequence. An initial Qp value Qp_0 is used to run the first-pass encoding. After encoding, the corresponding bit numbers and PSNR values of each frame i are collected as R_i and D_i.

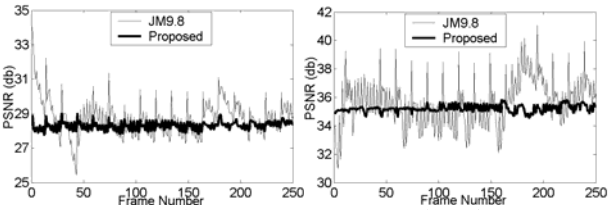
During the offline processing, the R-D models $PSNR_i(R)$ and $PSNR_i(Q)$ for each frame i are estimated first. The parameters a_i, b_i in function (9) and d_i in function (10) are set to constant values as 5, 2, 55 in I pictures, 7.6, 30, 52.4 in P pictures, and 10, 125, 52.2 in B pictures, based on our simulation results. The parameter B_i can be estimated by computing the PSNR of picture residuals. Then parameters A_i and c_i can be computed easily using R_i, D_i and Qp_0.

Since model parameters were estimated, the optimized target bit numbers and Qp values for each frame will be calculated by our bit allocation algorithm as follows:

1) Calculate the average distortion $\bar{D}$ with target average bit rate $\bar{R}$:
$$\bar{D} = \frac{1}{N}\sum_{i=0}^{N-1} PSNR_i(\bar{R})$$
where N is the total frame number.
2) Calculate the initial bit numbers allocation in frame i:
$$init\_bit_i = R_i(\bar{D})$$
where $R_i(D)$ is the inverse function of (9) in frame i.
3) The average tune bit number can be obtained by
$$tune\_bit = \frac{1}{N}(\sum_{i=0}^{N-1} init\_bit_i) - \bar{R}$$
4) Calculate the tune weight of each frame i:
$$Tune\_Weight_i = N \times [D_i'(init\_bit_i)]^{-1} \Big/ \sum_{i=0}^{N-1}[D_i'(init\_bit_i)]^{-1}$$
where $D_i'(init\_bit_i)$ is the derivative of function (9) at the bit rate *init_bit$_i$*, and $[D_i'(init\_bit_i)]^{-1} = R_i'(\bar{D})$ approximates the bit rate requirement of one unit distortion change at the distortion point $\bar{D}$ of frame i.
5) Then the target bit rate can be calculated by
$$target\_bit_i = init\_bit_i - tune\_rate \times Tune\_Weight_i$$
6) Calculate the average distortion $\bar{D}$ with target bit rate *target_bit$_i$*:
$$\bar{D} = \frac{1}{N}\sum_{i=0}^{N-1} PSNR_i(target\_bit_i)$$
7) Finally, the target Qsteps and Qps are calculated by:
$$Q_i = PSNR_i^{-1}(\bar{D})$$
$$Qp_i = round(6\log_2 Q_i + 4)$$
where $PSNR_i^{-1}(D)$ is the inverse function of (10) in frame i.

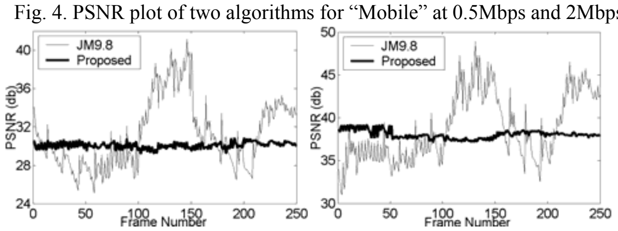
There are some differences between the target $\bar{R}$ and the average bit rate $\frac{1}{N}\sum_{i=0}^{N-1} target\_bit_i$. Usually the difference can be ignored. However, if needed, it can also be decreased by recursively calculating step 3 to step 5. The computation complexity of this bit allocation algorithm is *O(N)*.

Once we complete the offline processing, a group of Qp_i were generated for each frame i. Using these optimized Qps, the second-pass encoding can achieve consistent visual quality under the target bit rate. And the computation complexity of this two-pass VBR algorithm is *O(N)*.



(a) 0.5Mbps (b) 2Mbps

Fig. 4. PSNR plot of two algorithms for "Mobile" at 0.5Mbps and 2Mbps



(a) 0.5Mbps (b) 2Mbps

Fig. 5. PSNR plot of two algorithms for "New_seq" at 0.5Mbps and 2Mbps

4. EXPERIMENTAL RESULTS

In this Section, we describe experiments based on H.264/AVC standard to illustrate the effectiveness of the proposed algorithm. Similar results can be obtained using other video coding standards. We implement the proposed algorithms on the H.264 reference encoder software JM9.8, and compare the results with initial rate control algorithm. The experiments test two video sequences: "Mobile" and a new sequence "New_seq". Both of the sequences are chroma format 4:2:0, CIF, and 250 frames. The "New_seq" consists of segments from five sequences: "Mobile", "Football", "Foreman", "Flower", and "Paris", 50 frames each. The frame rate is 30 and GOP length is 15.

Fig. 4 shows PSNR comparison of "Mobile" between our algorithm and JM9.8 at the bit rate of 0.5 Mbps and 2.0Mbps. The PSNR curve of the JM algorithm has large variation at the beginning of the sequence, and then become a little smoother after 50 frames. However, the proposed algorithm has a much smoother curve during the whole sequence. Fig. 5 illustrates PSNR comparing with "New_seq", which simulates a frequently-changing-scene video sequence. Obviously, the proposed algorithm has a much smoother result than the JM.

Table. 1 provides the detailed results. Mean, variance, maximum, and minimum values of PSNR are also shown in the table. It is clearly that under same bit rate constraint, our algorithm achieves considerably smoothed PSNR value when compared to JM9.8, without notable PSNR reduction. In average, the variance of PSNR reduces greatly to less than 3% of the results by JM9.8. Especially for the "New_seq" sequence, the variance reduces to about 1%, which is 0.1286 versus 12.9712.

Table. 1.Performance of two algorithms

Sequence	Target rate (kbps)	Algorithm	Actual rate (kbps)	PSNR (db)			
				Mean	VAR	MAX	MIN
Mobile	500	JM9.8	502.74	28.85	1.1991	34.035	25.445
		Proposed	499.99	28.28	0.0724	29.039	27.749
	2000	JM9.8	2000.08	35.62	3.1160	41.059	30.993
		Proposed	2054.71	35.20	0.0712	35.867	34.561
New_seq	500	JM9.8	500.02	31.24	12.9712	41.178	25.055
		Proposed	496.79	30.06	0.1286	30.901	29.205
	2000	JM9.8	1999.68	38.78	15.4572	48.880	30.993
		Proposed	2018.94	38.04	0.2595	39.170	37.105

5. CONCLUSION

In this paper, we presented a new rate-distortion model for H.264. This model was developed from the statistical analysis of integer transform coefficients using GGD. Then, a novel two-pass VBR coding algorithm was proposed. Experimental results showed that the proposed algorithm was effective and efficient to reduce the PSNR variation and provide consistent visual quality, especially for frequently-changing-scene video sequences.

6. REFERENCES

[1] T. Wiegand, G.J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC Video Coding Standard", *IEEE Trans. Circuit Syst. Video Technology*, Vol. 13, No.7, pp. 560-576, Jul. 2003

[2] Y. Yu, J. Zhou, Y. Wang, and C. Chen, "A novel two-pass VBR coding algorithm for fixed-size storage application", *IEEE Trans. Circuit Syst. Video Technology*, Vol. 11, No.3, pp. 345-356, Mar. 2001

[3] H.B. Yin, X.Z. Fang, L. Chen, and J. Hou, "A practical consistent-quality two-pass VBR video coding algorithm for digital storage application", *IEEE Trans. Consumer Electronics*, Vol. 50, No. 4, pp. 1142-1150, Nov. 2004

[4] C. Que, G. Chen, and J. Liu, "An Efficient Two-Pass VBR Encoding Algorithm for H.264", *2006 International Conference on Communications, Circuits and Systems Proceedings*, Vol.1, pp. 118-122, Jun. 2006

[5] K. Sharifi, and A. Garcia, "Estimation of shape parameter for generalized Gaussian distribution in subband decomposition of video", *IEEE Trans. Circuits Syst. Video Technology*, Vol. 5, No. 1, pp. 52-56, Feb. 1995

[6] H.S. Malvar, A. Hallapuro, M. Karczewicz, and L. Kerofsky, "Low-complexity transform and quantization in H.264/AVC", *IEEE Trans. Circuits Syst. Video Technology*, Vol. 13, No. 7, pp. 598-603, Jul. 2003

[7] J. Sun, W. Gao, D. Zhao, and Q. Huang, "Statistical model, analysis and approximation of rate-distortion function in MPEG-4 FGS videos", *IEEE Trans. Circuits Syst. Video Technology*, Vol. 16, No. 4, pp. 535-539, Apr. 2006

[8] H. Gish, and J. Pierce, "Asymptotically efficient quantizing", *IEEE Trans. Inform. Theory*, Vol.14, No.5, pp.676-683, Sept. 1968