FRAME BIT ALLOCATION FOR H.264 USING CAUCHY-DISTRIBUTION BASED SOURCE MODELLING

Nejat Kamaci and Yucel Altunbasak

Center for Signal and Image Processing
Georgia Institute of Technology
Atlanta, GA 30332
email:{kamaci,yucel}@ece.gatech.edu

ABSTRACT

Based on the observation that a Cauchy density is more accurate in estimating the distribution of the AC coefficients than the traditional Laplacian density for H.264 video coders, a rate model with improved accuracy is derived. Using the new rate model, a frame bit-allocation algorithm for the rate control of an H.264 video coder is developed. Extensive analysis with carefully selected anchor video sequences demonstrates a 0.43 dB average PSNR improvement over the JM 8.4 rate control algorithm, and a 0.27 dB average PSNR improvement over the TM5-based bit-allocation algorithm that has recently been proposed for H.264 by Ma et al. The analysis also demonstrates 20% and 60% reductions in PSNR variation among the encoded pictures when compared to the JM 8.4 rate control algorithm and the TM5-based bit-allocation algorithm, respectively.

1. INTRODUCTION

Over the past few decades, transform-based compression for image and video sources has gained widespread popularity for visual information management, processing, and communications. As a result, several industry standards have been developed, such as MPEG-2 [1] and H.264 [2] for video coding. With all of these image and video processing methods, the image frame(s) is divided into nonoverlapping blocks, and a transformation is applied to the block before quantization and entropy coding. The two-dimensional discrete cosine transform (DCT) is the most common transform used in these methods[1].

The H.264/AVC (equivalently MPEG-4 Part 10) standard is one of the most advanced video coding standards that has been developed. H.264/AVC features a number of new technologies such as intra prediction, integer transform and multi-frame prediction, in addition to improvements on the existing technologies. In common with the earlier standards, the H.264/AVC does not explicitly define an encoder-decoder pair. Many functional parts of the encoder and the decoder are left open for optimization. One of these functional parts is the rate-control module that is responsible for controlling the output bit rate of the encoder.

1.1. Rate control for H.264/AVC

The output bit rate and video quality of a video encoder depend on several coding parameters such as the quantization scale (Q) and coding mode. In particular, choosing a large quantization scale reduces the resulting bit rate, while at the same time reducing the visual quality of the encoded video. In most applications, a predetermined constant output bit rate is desired. These applications are referred to as constant bit-rate (CBR) applications. The output bit rate of an encoder can be controlled by carefully selecting the quantization parameters for each coding block. This task is performed by the rate-control module. The goal of rate-control unit is to keep the output bit rate within constrained limits while achieving maximally uniform video quality.

For practical reasons, the rate-control problem is usually studied in three subproblems: (i) GOP bit allocation, (ii) picture bit allocation, and (iii) macroblock Q selection. GOP bit allocation involves selecting the number of bits to allocate to a GOP, which in the case of CBR rate-control, simply amounts to assigning a fixed number of bits per GOP. Picture bit allocation involves distributing the GOP budget among the picture frames, so as to achieve a maximal, uniform video quality. Macroblock Q selection involves tuning the Q parameter for each macroblock of a frame so that the the rate regulations are met and a uniform quality is achieved within the picture. As in the H.264 reference software, Q selection may also affect the motion estimation and compensation operations.

Many conventional rate-control algorithms use rate and distortion models for their operation. The performance of a rate-control algorithm greatly depends on its ability to estimate the rate and the distortion. In our earlier study we have demonstrated that the accuracy of modelling rate-distortion relation can be improved by using a Cauchy-distribution fit to the DCT coefficients, especially for the H.264 video coders [3]. The Cauchy distribution's heavy tails help estimating the actual statistical distribution of the transform coefficients more accurately for larger values. Consecutively, it helps improving the accuracy of the rate and distortion models. In this study, we demonstrate the benefit of using Cauchy-distribution based rate model in frame bit allocation for rate control purposes.

2. RATE MODELLING USING CAUCHY-DISTRIBUTION

The entropy of a uniformly quantized Cauchy source with parameter μ and quantization level Q is given in [3] by

$$\begin{aligned} H(\mu, Q) \quad &= \quad -\frac{2}{\pi}\xi\left(\frac{1}{2}, 2\mu, Q\right)\log_2 \xi\left(\frac{1}{2}, 2\mu, Q\right) \\ &\quad -\frac{2}{\pi}\sum_{i=1}^{\infty}\xi(i, \mu, Q)\log_2 \xi(i, \mu, Q), \end{aligned} \quad (1)$$

[1]The H.264 coder uses an integer transform that is a close approximation to the DCT.

where

$$\xi\left(i,\mu,Q\right)=\tan^{-1}\left(\frac{\mu Q}{\mu^2+\left(i^2-1/4\right)Q^2}\right).$$

This entropy functions based on the Cauchy is computable, provided that the density parameter μ is known. The parameter μ can be computed using the histogram of the AC coefficients.

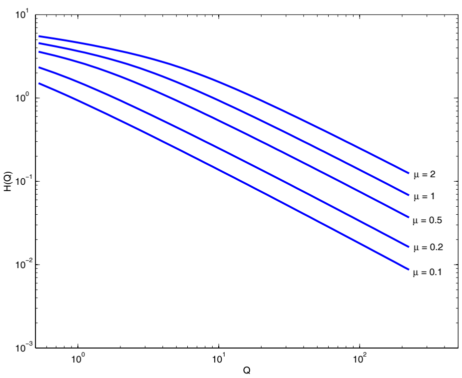


Fig. 1. Log-log plot of the entropy function given in Eq. (1) for five different values of μ.

Fig. 1 shows log-log plots of the entropy function given in Eq. 1 for five different μ values. The plots show that there is a nearly linear relation between $\ln(Q)$ and $\ln(H(Q))$, especially for smaller μ values and for $Q > 2$. Therefore we can approximate the expression given in Eq. 1 as

$$H\left(Q\right)\approx aQ^{-\alpha}, \tag{2}$$

where $a, \alpha > 0$ are parameters that depend on μ. These parameters can be calculated offline using Eq. (1).

3. FRAME BIT ALLOCATION FOR H.264

As part of the rate control problem of a video coder, the rate control module distributes a given GOP bit budget among the pictures of the GOP. The required bit budget for each picture frame that will result in a constant quality video output varies with the picture content and the picture coding type. The overall picture bit budget is shared by the encoding of the motion vectors, the quantized DCT coefficients, and the overhead bits.

We consider constant-bit-rate scenarios in which a constant number of bits is used within a GOP with a constant quantization parameter within a frame for each frame. Assume that we have a GOP size of G frames. Also assume that we are about to encode the k^{th} frame of the GOP, so that the previous $k-1$ frames have already been encoded. Assume that we have a bit budget of $R_{gop}(k)$ for the remaining frames of the GOP with indexes $(k, k+1, ..., G)$. We want to find the bit budget R_i for the i^{th} frame so that

$$\sum_{i=k}^{G} R_i = R_{gop}(k),$$

and

$$D_i \approx D_j, \qquad i \neq j,$$

where D_i is the distortion of the i^{th} frame. For H.264, selection of $Q_I = \nu \times Q_P = \nu \times Q_B$ with $\nu = 0.9$ roughly produces equal picture qualities for I,P and B pictures. Also for simplicity, we assume that the newly developed rate model of Eq. (2) is valid for the whole frame. Using the newly developed rate model given in Eq. (2), the target number of bits R_i for the i^{th} frame can be calculated as

$$R_i = \upsilon_i a_i a_k^{-\alpha_i/\alpha_k} R_k^{\alpha_i/\alpha_k} \left(\upsilon_i = \begin{cases} 0.9 & \text{if intra} \\ 1 & \text{else} \end{cases}\right),$$

where

$$\sum_{i=k}^{G} \upsilon_i a_i a_k^{-\alpha_i/\alpha_k} R_k^{\alpha_i/\alpha_k} = R_{gop}(k). \tag{3}$$

This requires computation of the parameters a_i, α_i of the rate-distortion relation given in Eq. (2) for frames $i = k, ..., G$. These parameters are not known before encoding, thus they need to be estimated. Estimation of these parameters will be discussed in Section 3.2.

Eq. (3) can be solved iteratively using Newton's method. Let

$$f\left(R_k\right) = \sum_{i=k}^{G} \upsilon_i a_i a_k^{-\alpha_i/\alpha_k} R_k^{\alpha_i/\alpha_k} - R_{gop}(k)$$

and let R_k^m be the value of R_k at the m^{th} step. The following iteration can be used to find R_k.

$$
\begin{aligned}
& R_k^0 = \frac{R_{gop}(k)}{G},\ \text{m} = 1. \\
& \text{while } \left|R_k^m - R_k^{m-1}\right| > \delta,\ \text{repeat:} \\
& \qquad R_k^m = R_k^{m-1} - \frac{f\left(R_k\right)}{f'\left(R_k\right)}; \\
& \qquad m++; \\
& R_k = R_k^m.
\end{aligned}
$$

The quantization parameter for the frame can be computed using Eq. (2):

$$Q_k = \left[\!\left[\left(\frac{R_k}{a_k}\right)^{-1/\alpha_k}\right]\!\right],$$

where $[\![.]\!]$ denotes rounding to the nearest possible quantization level.

3.1. Macroblock quantization adaptation

In this study, we consider a fixed quantization parameter within a frame, therefore quantization parameters of all macroblocks of a frame is fixed to the level computed by the frame bit allocation procedure.

3.2. Parameter handling and practical issues

In this paper, we focus on single-pass implementation scenarios in which we do not have prior information about the statistical properties of the video source. Hence, the model parameters (a, α) of the approximate rate model of Eq. (2) are not known and need to be estimated.

We propose the following approach: we consider separate estimators (a_I, a_P, a_B) and $(\alpha_I, \alpha_P, \alpha_B)$ for each of the picture types (I, P, B) for parameters a and α, respectively. This is because the compression efficiency hence the rate-distortion relation is different for each of the picture types, due to distinctions in coding methods. For each picture with index i in a GOP, parameters a_i, α_i of the picture are estimated using previously encoded pictures as

$$(a_i, \alpha_i) = \begin{cases} (a_I, \alpha_I) & \text{if } i \text{ is an I - type picture} \\ (a_P, \alpha_P) & \text{if } i \text{ is a P - type picture} \\ (a_B, \alpha_B) & \text{if } i \text{ is a B - type picture} \end{cases}.$$

As suggested in [6], we consider initialization of the frame bit allocation algorithm by choosing a frame quantization parameter for each frame type as follows: Define B_{pp} as the average number of bits targeted per a video frame pixel. It is calculated as

$$B_{pp} = \frac{R}{F \times N_p},$$

where R is the target bit rate, F is the frame rate, and N_p is the number of pixels per frame (*e.g.*, for a $4 : 2 : 0$ format QCIF sequence, $N_p = 176 \times 144 \times 1.5$). For the given bit rate R, the initial quantization parameter for the first intra picture, Q_0^I, is decided as

$$Q_0^I = \begin{cases} 40 & B_{pp} < 0.05 \\ 30 & 0.05 \le B_{pp} \le 0.1 \\ 20 & B_{pp} > 0.1 \end{cases}.$$

Consequently, $Q_0^P = Q_0^B = Q_0^I + 1$. Using the first set of pictures, we estimate the model parameters as follows: after encoding the first intra picture, we collect the DCT statistics, calculate output bits (R_I), and set

$$\alpha_I = \begin{cases} 0.75 & \text{if } \mu < 1.0 \\ 0.85 & \text{if } \mu > 2.0 \\ 0.8 & \text{else} \end{cases}, \text{and} \quad a_I = \frac{R_I}{(Q_0^I)^{-\alpha_I}}.$$

Also, after encoding first P- and B- pictures, we calculate the output bits (R_P, R_B) and set

$$\alpha_P = \begin{cases} 1.2 & \text{if } \frac{R_{P,act}}{N_p} > 0.1 \\ 1.6 & \text{if } \frac{R_{P,act}}{N_p} < 0.05 \\ 1.4 & \text{else} \end{cases},$$

$$\alpha_B = \begin{cases} 1.6 & \text{if } \frac{R_{B,act}}{N_B} > 0.1 \\ 2.0 & \text{if } \frac{R_{B,act}}{N_B} < 0.05 \\ 1.8 & \text{else} \end{cases}, \text{ and}$$

$$(a_P, a_I) = \left(\frac{R_P}{(Q_0^I)^{-\alpha_I}}, \frac{R_B}{(Q_0^B)^{-\alpha_B}} \right).$$

For the remaining pictures in the video sequence, the model parameters are updated as follows: (i) α is fixed, and (ii) parameter a for picture type x ($x \in (I, P, B)$) is updated using

$$a_x = \delta \times a_x + (1 - \delta) \frac{R_x}{(Q)^{-\alpha_x}},$$

where δ is a forgetting factor. We choose $\delta = 0.5$ in our simulations.

3.3. Proposed frame bit allocation algorithm

For a given bit rate target R and a frame rate F, the GOP bit budget (C_{gop}) is given as $C_{gop} = R/F$. The bit budget for frame i is calculated as follows:

1. Initialization: Calculate Q_0^I, Q_0^P and Q_0^B. Initialize the model parameters as described in the parameter handling section.
2. Start of a GOP: Set $R_{gop}(1) = C_{gop}$, where C_{gop} is the constant target GOP rate.
3. For $k = 1$ to G repeat:
   (a) Solve for R_k using
   $$\sum_{i=k}^{G} \upsilon_i a_i a_k^{-\alpha_i/\alpha_k} R_k^{\alpha_i/\alpha_k} = R_{gop}(k).$$
   (b) Solve for Q_k using
   $$Q_k = \left[\left[\left(\frac{R_k}{a_k} \right)^{-1/\alpha_i} \right]\right].$$
   (c) Encode frame, and calculate actual output bits, S_k.
   (d) Update $R_{gop}(k+1) = R_{gop}(k) - S_k$.
4. Update the model parameter
   $$a_k = \delta a_k + (1 - \delta) \frac{S_k}{(Q_k)^{-\alpha_k}}$$
5. If end of the sequence reached, stop.
   Else go to step 2.

4. APPLICATION RESULTS

For evaluating its performance, we implemented the proposed frame bit-allocation algorithm on the joint video team (JVT) H.264 reference encoder software. For comparison, we considered two other methods: (i) the rate control algorithm of JM 8.4, and (ii) the improved TM5-based frame bit-allocation algorithm proposed for H.264 in [6].

In the tests, we used an [IPPP...] GOP structure of size 12. We used the first 120 frames of each sequence. The H.264 encoder was configured to have two reference frames for inter motion search, context-based adaptive binary coding (CABAC) for symbol coding, rate-distortion optimized mode decisions, and full search motion estimation with a search range of 16. The PSNR values are measured on the luminance component only.

Table 1 summarizes the encoding results. Average PSNR values, and the PSNR variation between frames are also shown in the table. The rate control method using the proposed frame bit-allocation algorithm achieves an average of 0.27 dB PSNR gain over the one that uses the TM5-based frame bit allocation algorithm proposed in [6] and an average of 0.43 dB PSNR gain over JM 8.4 rate control method [5]. In addition, the proposed algorithm achieves considerably reduced PSNR variation between frames on the average when compared to the other algorithms. The proposed algorithm achieves 20% and 60% reductions in PSNR variation among the encoded pictures when compared to the JM 8.4 rate control algorithm and the TM5-based bit-allocation algorithm, respectively.

For further evaluation, Fig. 2 shows the PSNR versus frame plots for each video sequence. The proposed algorithm shows superior performance by achieving a consistent video quality throughout the sequence for all video sequences.

Sequence	Method	Output rate (Kbps)	Average PSNR (dB)	PSNR variance (dB)
TEMPETE (CIF)	(a)	1505.34	35.48	4.40
	(b)	1502.06	35.45	1.43
	(c)	1500.74	35.78	0.68
PARIS (CIF)	(a)	700.18	35.24	3.95
	(b)	704.96	35.10	0.54
	(c)	699.99	35.67	0.66
FOREMAN (QCIF)	(a)	100.03	33.58	3.25
	(b)	100.51	33.86	1.80
	(c)	99.66	33.92	0.72
AKIYO (QCIF)	(a)	59.87	37.86	2.37
	(b)	60.36	38.08	1.52
	(c)	59.87	38.35	0.76
NEWS (QCIF)	(a)	99.91	35.59	3.79
	(b)	100.33	35.67	1.52
	(c)	99.71	35.87	0.86
SILENT (QCIF)	(a)	80.07	33.40	1.27
	(b)	80.48	33.42	0.92
	(c)	80.01	34.12	1.73

Table 1. Performances of three algorithms: (a) TM5-based, (b) JM 8.4, and (c) proposed Cauchy-distribution based algorithms in terms of output rate, average psnr, and psnr variation.

5. SUMMARY AND CONCLUSION

We demonstrated the practicality and effectiveness of the Cauchy-domain rate and distortion models by using them in a frame bit-allocation problem. The proposed frame bit-allocation algorithm is capable of achieving an average of 0.43 dB PSNR gain compared to JM 8.4 rate control algorithm and an average of 0.27 dB PSNR gain compared to an improved TM5-based frame bit-allocation algorithm proposed for H.264 video coder. Furthermore, the proposed algorithm helps reducing the PSNR variation among the frames when compared to the two algorithms. We attribute the resulting improvements with the proposed frame bit allocation algorithm to more accurate and robust modelling by the use of the Cauchy-distribution for estimating the transform coefficients. Cauchy-distribution's extreme heavy tails have the effect of making the rate model always be on guard for an outlier. In effect, compared with other rate models, it would be less likely to overreact to local changes in activity. The proposed frame bit allocation method can be extended to a complete rate control algorithm by incorporating a macroblock layer quantization adaptation method.

6. REFERENCES

[1] "The MPEG-2 international standard," ISO/IEC, Reference number ISO/IEC 13818-2, 1996.

[2] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. on Circuits and Systems for Video Technology,* vol. 13, pp. 560-576, July 2003.

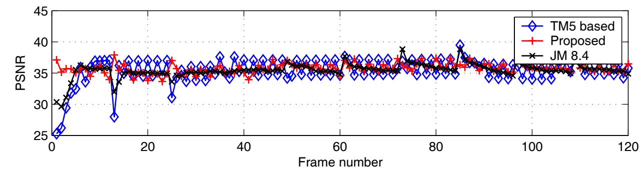
[3] "An analysis of the DCT coefficient distribution with the H.264 video coder," *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing* , vol. 3, May 2004 pp. 177 - 180

[4] J. Ribas-Corbera and S. Lei, "Rate control in DCT video coding for low-delay communications," *IEEE Trans. on Circuits and Systems for Video Technology,* vol. 9, pp. 172-185, Feb. 1999.

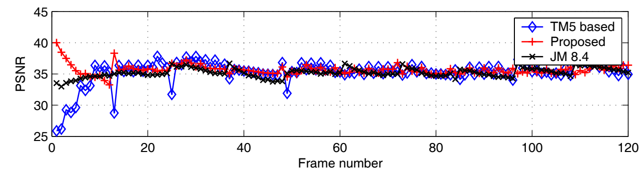
[5] Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, Joint Model Reference Software Version 8.4.

[6] S. Ma, W. Gao, F. Wu and Y. Lu, "Rate Control for advance video coding (AVC) standard," *Proceedings of the 2003 International Conference on Image Processing, ICIP '03,* Sept. 2003, vol. 3, pp. 793-796.

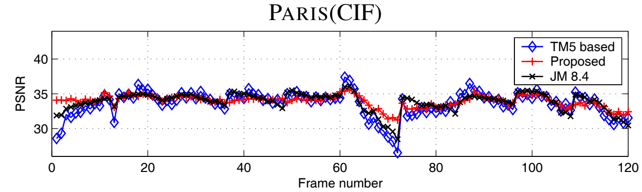
7. FIGURES



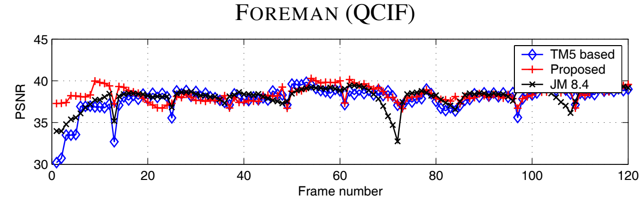
TEMPETE (CIF)



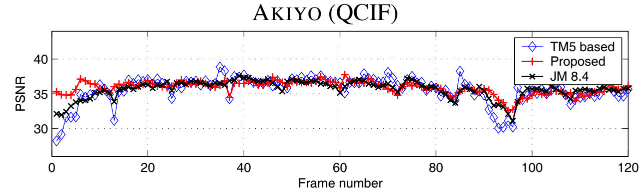
PARIS(CIF)



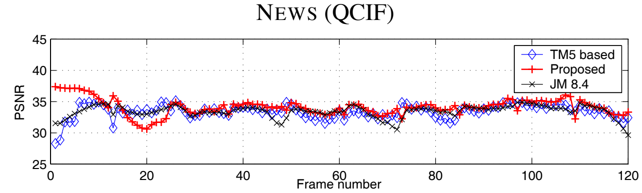
FOREMAN (QCIF)



AKIYO (QCIF)



NEWS (QCIF)



SILENT (QCIF)

Fig. 2. psnr vs. frame for three rate control algorithms.