MATHEMATICAL ANALYSIS OF MPEG COMPRESSION CAPABILITY AND ITS APPLICATION TO RATE CONTROL

Jiro Katto and Mutsumi Ohta

Information Technology Research Laboratories, NEC Corporation
1-1, Miyazaki 4-chome, Miyamae-ku,
Kawasaki-shi, Kanagawa 216, JAPAN
E-Mail: katto@DSP.CL.nec.co.jp

ABSTRACT

This paper presents mathematical frameworks on temporal predictive processing in the MPEG video compression standard. Firstly, a coding gain is derived based on traditional prediction theories. The optimum ordering of three different picture types (I,P,B-pictures) is clarified according to image source characteristics. Secondly, a novel framework on the target bit assignment is presented with some experimental backgrounds. The solution consists of simple formulae, but brings drastic SNR gains to the conventional TM5 algorithm.

1. INTRODUCTION

Bidirectional prediction is one of the factors characterizing the MPEG video compression standard [1, 2]. Coding efficiency is highly improved for some image sequences when the bidirectional prediction is utilized in an adequate manner. However, known theories only indicate that interpolative prediction leads to lower prediction errors than extrapolative prediction [3], and total balance of the prediction mode arrangement has been never considered. The picture ordering has been empirically defined in the MPEG standard without sufficient theoretical backgrounds.

Rate control algorithms have been introduced heuristically due to lack of theoretical supports. Test model 5 (TM5) developed for the MPEG2 standard [4] provides an efficient algorithm with three steps: target bit assignment (step 1), feedback control (step 2) and modulation (step 3). A lot of papers have been published to improve the rate control behavior. However, they are dedicated to the issues related to step 2 and step 3 parts [5]-[9], and the step 1 part has not been investigated enough. There is a paper discussing bit assignment issues [10], but its solution seems to be too complicated for practical use.

This paper presents two mathematical frameworks for these problems. In Section 2, a coding gain is provided based on orthodox prediction theories. A parameter reflecting persistent tendency of interframe correlations is introduced, and the optimum picture ordering is indicated. In Section 3, an improved algorithm for target bit assignment is proposed by exploiting experimental rate distortion functions. The solution is a TM5 extension but provides drastic SNR gains.

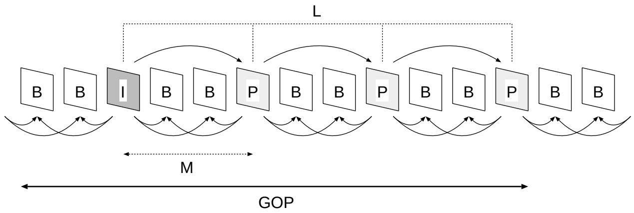

Figure 1: A GOP structure in the MPEG standard

2. A CODING GAIN FRAMEWORK INDICATING OPTIMUM PICTURE ORDERING

2.1. Periodic Picture Ordering in the MPEG Standard

The MPEG standard has three picture types from the viewpoint of their temporal processing; intra-coded picture (I-picture), predictive-coded picture (P-picture) and bidirectionally predictive-coded picture (B-picture). A group of pictures (GOP) is then defined as shown in Figure 1. We can represent this GOP structure by

$$B^{M-1}I(B^{M-1}P)^{L-1} \tag{1}$$

where M is a distance between "core pictures" (I/P-pictures) and L is the number of core pictures in the GOP. $N = LM$ is then equal to the total number of pictures in the GOP (i.e. an interval between I-pictures).

2.2. Derivation of a Coding Gain

We assume following relationships:

- Rate distortion function is given by

$$\sigma_q^2 = \epsilon^2 2^{-2R} \sigma_x^2, \tag{2}$$

  where σ_x^2 is an input source variance, σ_q^2 is a quantization error variance, R is allocated bits and ϵ^2 is a quantization performance factor [11].

- Closed-loop prediction guarantees $\sigma_{r,t}^2 = \sigma_{q,t}^2$ for each t-picture (t=I,P,B), where $\sigma_{r,t}^2$ means a reconstruction error variance.

- Prediction error variances are given by

$$2 \cdot \Big(1 - cor(M)\Big) \cdot \sigma_x^2 \tag{3}$$

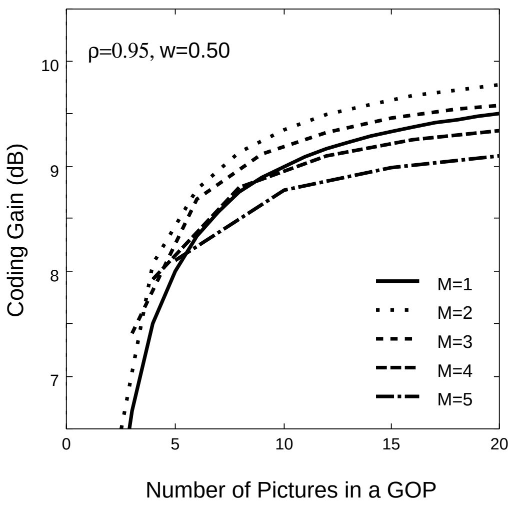

Figure 2: Coding gains for input sources with typical correlation continuity (w=0.5).

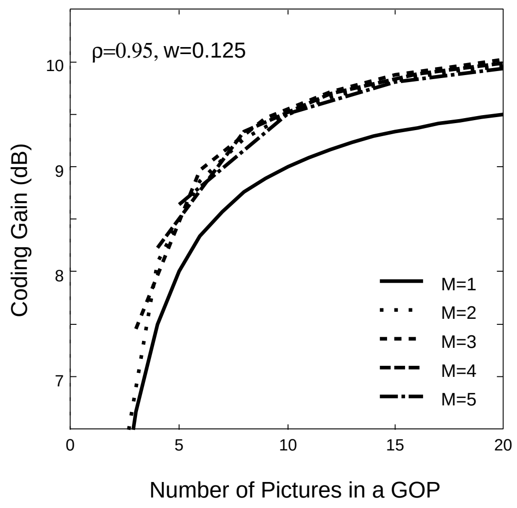

Figure 3: Coding gains for input sources with highly continuous correlations (w=0.125).

for P-picture, and

$$\left(\frac{3}{2} + \frac{1}{2} \cdot cor(M) - cor(j) - cor(M-j)\right) \cdot \sigma_x^2 \quad (4)$$

for B-picture, where j is a distance from the past core picture and $cor(k)$ represents a correlation between k-frame apart pictures.

Based on these equations, an optimum bit allocation problem is formulated as follows: minimizing the average reconstruction error variance

$$\sigma_r^2 \ = \ \frac{1}{LM}\left[\sigma_{r,I}^2 + (L-1) \cdot \sigma_{r,P}^2 + L \cdot \sum_{j=1}^{M-1} \sigma_{r,B}(j)^2\right] \quad (5)$$

on the constant rate constraint

$$R_I + (L-1) \cdot R_P + L(M-1) \cdot R_B \ = \ const. \quad (6)$$

In Eq. (6, R_t represents allocated bits to t-picture.

By applying the Lagrange multiplier method, a coding gain of the GOP structure is defined by

$$G_{IPB} \ = \ \frac{1}{P(M)^{\frac{L-1}{LM}} \cdot \left[\frac{B(M)}{M-1}\right]^{\frac{M-1}{M}}}, \quad (7)$$

where $P(M)$ and $B(M)$ are

$$P(M) \ = \ 2 \cdot (1 - cor(M))$$
$$B(M) \ = \ \sum_{j=1}^{M-1}\left(\frac{3}{2} + \frac{1}{2} \cdot cor(M) - cor(j) - cor(M-j)\right), \quad (8)$$

respectively.

2.3. Comparison of Coding Gains

We introduce an input source model of which $cor(k)$ is given by

$$cor(k) \ = \ \rho^{k^w}, \quad (9)$$

where ρ is a correlation coefficient between neighboring pictures and w is a weighting factor. When $w = 1$, this is a traditional AR(1) model. The objective of w is to reflect persistent tendency of interframe correlations, which is suggested by simulation results using real image sequences. Typical values of w are around 0.25 $\sim$ 0.5. It sometimes takes very small values (< 0.125) in such a case of Mobile & Calendar. This means that correlation still remains between distant pictures.

Applying Eq. (9) to Eq. (7), coding gains are calculated for various picture ordering patterns. Figure 2 shows an example of $\rho = 0.95$ and w=0.50. This is corresponding to the case of ordinary image sequences. Figure 3 depicts another example (w=0.125) corresponding to image sources with highly continuous correlations. These figures point out that

- Adequate insertion of B-pictures brings SNR gains to the M=1 case (no B-picture),
- $M = 2$ or 3 will be the best choice for usual image sequences, and
- $M > 3$ remains useful for highly correlated image sequences.

Coding simulations are carried out using six ITU-R standard pictures. The results are almost consistent with theoretical evaluations above. For example, Mobile & Calendar sequences, which have very high correlations one another (w is very small), shows that larger M values brings higher SNR gains. The other ordinary sequences, which have higher w values, have a peak at $M = 2$ or 3.

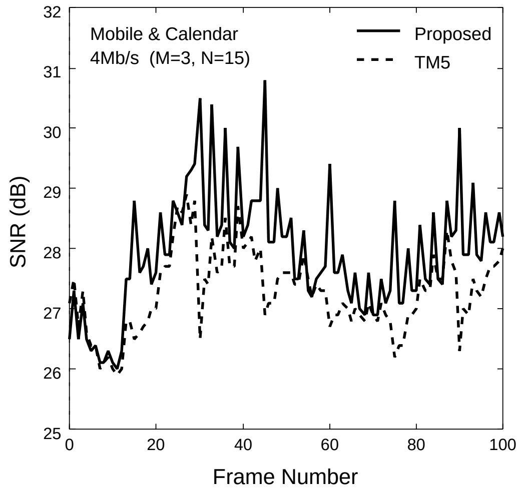

Figure 4: Comparion of SNR values per frame

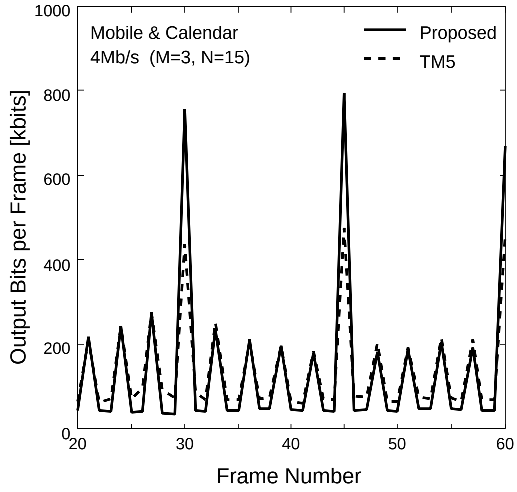

Figure 5: Comparison of generated bits per frame

3. OPTIMUM TARGET BIT ASSIGNMENT

3.1. A Motivation

A disadvantage of the previous framework is fragility of its bit assignment applications (R_B sometimes takes a negative value). We then represent the rate distortion function by

$$\log R \;=\; a \cdot \log Q + b \tag{10}$$

instead of Eq. (2), where Q is a quantization step size and a and b are constants. This equation is derived based on experimental results using real image sequences for the development of rate control algorithm [12, 13]. This is equivalent to

$$Q \cdot R^{\alpha} \;=\; X \tag{11}$$

where X and α are constants depending on image source characteristics. The complexity measure in TM5 is corresponding to the case of α=1.

3.2. Bit Assignment Formulation

Based on the rate distortion function in Eq. (11), another optimum bit allocation problem is formulated as follows: minimizing the average of the m-th order quantization step sizes

$$\frac{N_I \cdot Q_I^m + N_P \cdot Q_P^m + N_B \cdot Q_B^m}{N_I + N_P + N_B} \tag{12}$$

on the same bit rate constraint

$$N_I R_I + N_P R_P + N_B R_B \;=\; const, \tag{13}$$

where N_t is the number of remaining t-pictures in a GOP. In this framework, we assume that minimization of the m-th order quantization step sizes results in SNR maximization.

Applying the Lagrange multiplier method, the solution is given by

$$R_I \;=\; \frac{R}{N_I + N_P\left(\frac{X_P}{X_I}\right)^{\frac{m}{1+m\alpha}} + N_B\left(\frac{X_B}{X_I}\right)^{\frac{m}{1+m\alpha}}}$$

$$R_P \;=\; \frac{R}{N_P + N_B\left(\frac{X_B}{X_P}\right)^{\frac{m}{1+m\alpha}}} \tag{14}$$

$$R_B \;=\; \frac{R}{N_B + N_P\left(\frac{X_P}{X_B}\right)^{\frac{m}{1+m\alpha}}}.$$

This solution presents target bit assignment strategy to each picture, which don't suffer from the the fragility problem occurred in the previous framework. In addition,

$$Q_{I,opt} : Q_{P,opt} : Q_{B,opt} \;=\; X_I^{-\frac{1}{m\alpha+1}} : X_P^{-\frac{1}{m\alpha+1}} : X_B^{-\frac{1}{m\alpha+1}} \tag{15}$$

is suggested on the optimum quantization step sizes. These equations support the fact of utilizing different step sizes to each pictures.

3.3. Relationship to TM5

Target bit assignment algorithm in TM5 (called step 1) is connected to the proposed formulation by setting

$$\alpha \;=\; 1,\; K_P \;=\; \left(\frac{X_I}{X_P}\right)^{\frac{1}{m+1}},\; K_B \;=\; \left(\frac{X_I}{X_B}\right)^{\frac{1}{m+1}} \tag{16}$$

where K_P and K_B are called universal constants whose recommended values are 1.0 and 1.4, respectively [4]. In other words, the proposed strategy means "adaptive re-assignment of K_P and K_B according to image source characteristics."

3.4. Experiments

As an auxiliary experiment, optimization of two parameters, m and α, is investigated. The results show that $\alpha \;=\; 1.2 \sim 1.3$ and $1/(m+1) \;=\; 0.6 \sim 1.2$ lead to the highest SNR values.

Taking these values into account, MPEG2 coding simulations are carried out. Figure 4 demonstrate an SNR comparison result between the proposed algorithm and TM5 for Mobile & Calendar. Figure 5 shows the corresponding result of their generated bit rates per frame. These figures indicate that the proposed algorithm brings about higher SNR gains than TM5 mainly by assigning

Table 1: Summary of the presented formulations

	1st (Coding gain)	2nd (Target bit assignment)
R-D function	$\sigma_q^2 \;=\; \epsilon^2 2^{-2R} \sigma_x^2$	$Q \cdot R^{\alpha} \;=\; X$
Cost function	$\dfrac{\sigma_{r,I}^2 + (L-1)\cdot\sigma_{r,P}^2 + L(M-1)\cdot\sigma_{r,B}^2}{LM}$	$\dfrac{N_I \cdot Q_I^m + N_P \cdot Q_P^m + N_B \cdot Q_B^m}{N_I + N_P + N_B}$
Constraint	$R_I + (L-1)R_P + L(M-1)R_B \;=\; LM \cdot R$	$N_I R_I + N_P R_P + N_B R_B \;=\; R$
Coding gain	$G_{IPB} \;=\; \dfrac{1}{P(M)^{\frac{L-1}{LM}} \cdot \left[\frac{B(M)}{M-1}\right]^{\frac{M-1}{M}}}$	(*not given in an explicit form*)
Bit allocation	$R_I \;=\; R - \dfrac{L-1}{2LM}\log_2 P(M) - \dfrac{M-1}{2M}\log_2 B(M)$ $R_P \;=\; R + \dfrac{L(M-1)+1}{2LM}\log_2 P(M) - \dfrac{M-1}{2M}\log_2 B(M)$ $R_B \;=\; R - \dfrac{L-1}{2LM}\log_2 P(M) + \dfrac{M+1}{2M}\log_2 B(M)$ (*fragile for practical use*)	$R_I \;=\; \dfrac{R}{1 + N_P\left(\frac{X_P}{X_I}\right)^{\frac{m}{1+m\alpha}} + N_B\left(\frac{X_B}{X_I}\right)^{\frac{m}{1+m\alpha}}}$ $R_P \;=\; \dfrac{R}{N_P + N_B\left(\frac{X_B}{X_P}\right)^{\frac{m}{1+m\alpha}}}$ $R_B \;=\; \dfrac{R}{N_B + N_P\left(\frac{X_P}{X_B}\right)^{\frac{m}{1+m\alpha}}}$ (*including TM5*)

more bits to I/P-pictures. This bit assignment strategy is not almighty, and causes annoying distortions in B-pictures for some image sequences. However, it should be noticed that the proposed algorithm do this adaptively and automatically according to image source characteristics with the help of theoretical backgrounds. Other image sequences are also efficiently compressed; improvements by 0.40 dB in Table Tennis and by 0.38 dB in Flower Garden are achieved, for example.

4. CONCLUSIONS

This paper presents two mathematical frameworks for the temporal prediction strategy in the MPEG video compression standard. Table 1 summarizes main formulations presented in this manuscript. They convince us of effectiveness of the bidirectional prediction and make way to develop MPEG compression capability to the full extent.

5. ACKNOWLEDGMENT

The authors thank to Dr. T. Nishitani for his helpful suggestions.

6. REFERENCES

[1] IS 11172-2: "Information technology - Coding of moving pictures and associated audio for digital storage media at up to about 1.5Mbit/s: video part," (1993).

[2] IS 13818-2: "Information technology - Generic coding of moving pictures and associated audio," (1995).

[3] A. K. Jain: "Image coding via a nearest neighbors image model," *IEEE Trans. on Commun.*, pp.318-330 (1975).

[4] ISO/IEC-JTC1/SC29/WG11: "Test Model 5 (Drafts)," MPEG93/N0400 (1993).

[5] H. Watanabe and Y. Ito: "Mathematical analysis of rate control used in MPEG2 video coding," *PCS'94*, 3-03 (1994).

[6] M. Banic et al.: "MPEG-2 optimized video coder," *PCS'94*, 3-04 (1994).

[7] J. Choi and D. Park: "A stable feedback control of the buffer state using the controlled Lagrange multiplier method," *IEEE Trans. on Image Processing*, pp.546-558 (1994).

[8] G. Keesman et al.: "Bit-rate control for MPEG encoders," *Signal Processing: Image Communication*, pp.545-560 (1995)

[9] L. Wang: "Rate control for MPEG video coding," *SPIE VCIP'95*, pp.53-64 (1995)

[10] K. Ramchandran et al.: "Bit allocation for dependent quantization with applications to multiresolution and MPEG video coders," *IEEE Trans. on Image Processing*, pp.533-545 (1994).

[11] N. S. Jayant and P. Noll: "Digital coding of waveforms," *Englewood Cliffs NJ*, Prentice-Hall (1984).

[12] S. Sakazawa et al: "A study on coding control for multi-encoder system," PCSJ'94, 7-14 (1994, in Japanese).

[13] E. Murata et al: "Study of the relationship between quantizer step size and coded bit amount for video coding," *1994 IEICE Spring Conf.*, D-190 (1994, in Japanese).