A TRANSFORM VIDEO CODER SOURCE MODEL AND ITS APPLICATION

Jiann-Jone Chen and Hsueh-Ming Hang

Depart. of Electronics Eng.
National Chiao-Tung University,
1001, University Rd., Hsin-Chu 300, Taiwan, ROC
Fax:886-35-723238; Email:hmhang@cc.nctu.edu.tw

ABSTRACT

A source model describes the relationship between the bits, distortion, and quantization step sizes of a large class of block-transform video coder is proposed. This model is derived from rate-distortion theory, and verified by real images. It enables us to predict the bits needed to encode a picture for a given distortion or to adjust the quantization scales of a coder for a given bit rate. Based on this derived model, a variable frame rate coding algorithm is developed. It can be used to control the frame rate of a coder to ensure a minimum picture quality of every frame. Simulation results indicate that improved performance is obtained by using this variable frame rate coding scheme when compared to the simple approach of varying the quantization scale linearly proportional to the encoder buffer level.

1. INTRODUCTION

Transform coding transmission is very popular in image compression. It is one of the most important elements in the international visual communication standards [1]-[3]. In order to achieve the best tradeoff between picture quality and bit rate for a picture or a picture sequence, it is very desirable to have a source model for this type of coders. Essentially, we like to predict the bits needed to encode a picture under a given distortion or to estimate the quantization step size that will produce a preselected bit rate.

Several basic elements in our model have already exist in the literature; however, to our knowledge, they have not been put together to form a complete yet compact model for block-transform coders. In this paper, we first briefly review the known results in rate distortion theory and then derive our source model by combining these results together with our own extensions. In Section 3, the real picture characteristics and non-ideal factors in a practical coder are discussed. In Section 4, a variable frame rate (VFR) coding algorithm is developed based on this proposed model. The simulation results for VFR coding are discribed in Section 5. Section 6 concludes the paper.

This work was in part supported by the National Science Council of ROC under grant NSC83-0404-E-009-021

2. SOURCE MODEL

In this section, the 1-D (dimensional) discrete signal properties relevant to our subject are first reviewed, and then these properties are extended to the 2-D signals.

2.1. Stationary Gaussian Process

The well-known rate distortion function of a discrete stationary Gaussian process $\{x(n)\}$ under the mean square distortion criterion can be found in many reference ([4], [5]). Assuming a uniform sampling grid, the rate distortion function can be approximated by the following discrete versions:

$$R_\theta = \frac{1}{2L} \sum_{\omega_i \in A_1} \log_2 \frac{\Phi(\omega_i)}{\theta} \tag{1}$$

$$D(R_\theta) = \frac{1}{L} \sum_{\omega_i \in A_1} \theta + \frac{1}{L} \sum_{\omega_i \in B_1} \Phi(\omega_i), \tag{2}$$

where $\Phi(\omega_i)$ is the power spectrum density function of $x(n)$, and

$$\begin{cases} Region\, A_1: & \{\, i \in \{0,1,...,L-1\} \text{ and } \Phi(\omega_i) \geq \theta\} \\ Region\, B_1: & \{0,1,...,L-1\} \;-\; A_2\ , \end{cases}$$

L is the number of samples in a data block, and $\omega_i = i \cdot \frac{2\pi}{L}, i = 0,1,...,L-1$.

A case of interest is that at low distortion when $A_1 = (-\pi, \pi]$ (or B_1 is empty), eqn. (2) becomes $D = \theta$. And thus, eqn. (1) becomes

$$e^{L\,\alpha \cdot R(D)} = \prod_{i=0}^{L-1} \frac{\Phi(\omega_i)}{D}. \tag{3}$$

0-8186-6950-0/94 $4.00 © 1994 IEEE

Essentially we are approximating a joint Gaussian source by multiple i.i.d. Gaussian sources. The major approximation errors, if measured by percentage over the signal component magnitude, appear at the high frequency components whose correlations in spatial domain are at distances close to L.

Assuming the 2-D signals we are dealing with are separable in the horizontal and the vertical directions, then all the above properties can readily be extended to the 2-D signals without significant modifications. Although the Karthunen-Loève(K-L) transform is usually recognized as the optimal transform for decorrelating the data, for practical purpose, the separable DCT (Discrete Cosine Transform) that requires much less computations seems to be adequate for most of the natural pictures and thus is widely used [1]-[3],[7],[8].

2.2. Quantization

Scalar quantizer is often used in real systems not only because of its simplicity but also because of its adaptability to the local pictorial data. A uniform mid-tread quantizer (in which zero is a reconstruction level) is often used in a practical coding system. The behavior of such a quantizer can be analyzed for inputs with known probability density distribution.

At high bit rates (small distortion), the bits (b) versus distortion (D) relation of an entropy-coded uniform quantizer for a zero-mean i.i.d. source $X(\cdot)$ can be approximated by the following formulas [6]-[8]:

$$b(D) = \frac{1}{\alpha}\log_e\left(\epsilon^2 \cdot \frac{\sigma_X^2}{D}\right) = \frac{1}{\alpha \log_2 e}\log_2\left(\epsilon^2 \cdot \frac{\sigma_X^2}{D}\right), \tag{4}$$

and

$$D(b) = \frac{\Delta^2}{\beta}, \tag{5}$$

where β is 12 and α is 1.386 ($= 2/\log_2 e$) for uniform, Gaussian, and Laplacian distributions, ϵ^2 is source dependent and is around 1 for uniform distribution, 1.4 for Gaussian, and 1.2 for Laplacian, and σ_X^2 is the signal variance. Combining eqns. (4) and (5) we obtain

$$b(\Delta) = \frac{1}{\alpha}\log_e\left(\epsilon^2 \cdot \beta \cdot \frac{\sigma_X^2}{\Delta^2}\right) = \frac{1}{\alpha}\log_e\left(\gamma \cdot \frac{\sigma_X^2}{\Delta^2}\right), \tag{6}$$

where $\gamma = \epsilon^2 \cdot \beta$. This gives us a more direct relation between b and Δ.

This arrangement, a uniform quantizer followed by an ideal entropy coder, is close to the optimum entropy-constrained nonuniform quantizer [7]. In a real system, the ideal entropy coder is typically replaced by a Variable-Length-Coder (VLC), a simplified version of Huffman code [9]. The bits produced by a VLC, $\bar{b}$ may be approximated by $s_{VLC} \cdot b$, where b is the ideal entropy bits of the quantizer outputs, and s_{VLC} is a scaling factor greater than 1. Under this assumption, eqn (6) can still be used for a practical scalar quantizer with a modified value of α.

2.3. Practical transform coder

A practical image transform coder, such as the DCT coders used in [1]-[3], can be represented by the general block diagram in Fig. 1. The data model used in Fig. 1 simply rearranges the transform coefficients in a zigzag scan order. This zigzag ordering affects the statistics of the symbols used in the source model.

Assuming the probability distribution of the L frequency components is either uniform, Gaussian, or Laplacian, and b_i is the bits of the i^{th} entropy-coded quantized coefficient, the total average bits of such a source is

$$\bar{b}(\bar{D}) = \frac{1}{L\,\alpha}\log_e\left[\prod_{i=0}^{L-1}\left(\epsilon_i^2 \cdot \frac{\sigma_i^2}{D_i}\right)\right], \tag{7}$$

where D_i, σ_i, and ϵ_i^2 are the distortion, the variance, and the ϵ^2 parameter associated with the i^{th} component. Since $D_i = \Delta_i^2/\beta_i$ (eqn (5)),

$$\bar{D} = \frac{1}{L}\sum_{i=0}^{L-1} D_i = \frac{1}{L}\sum_{i=0}^{L-1}\frac{\Delta_i^2}{\beta_i}, \tag{8}$$

where Δ_i is the quantization step size of the i^{th} component, and β_i is the β parameter associated with that component.

Due to the frequency-dependent noise visibility of human perception [8], bits assigned to various frequency components should be adjusted according to the perceptual threshold. In [2] and [3], the quantization step sizes of transform coefficients are made of two components: q_s, a quantization scaling factor for the entire picture block, and $\{W_i, i = 0, \ldots L-1\}$, a weighting matrix whose elements are used as multiplication factors to produce the true step sizes in quantization. In other words, $\Delta_i = q_s \cdot W_i$.

Therefore, eqns (7) and (8) become

$$\bar{b}(\bar{D}) = \frac{1}{\alpha}\log_e\left\{\frac{1}{q_s^2}\left[\prod_{i=0}^{L-1}\left(\frac{\beta_i \cdot \epsilon_i^2 \cdot \sigma_i^2}{W_i^2}\right)\right]^{1/L}\right\} \tag{9}$$

and

$$\bar{D} = \frac{q_s^2}{L}\sum_{i=0}^{L-1}\frac{W_i^2}{\beta_i}. \tag{10}$$

The bits and distortion behavior of such a transform coder is thus described by eqns (9) and (10).

In reality, some frequency components may have an *effective variance* ($=\epsilon_i^2 \cdot \sigma_i^2$) less than the weighted distortion, $W_i^2 \cdot q_s^2/\beta_i$, at that frequency. We then need to go back to eqns (1) and (2), and modify eqns (9) and (10) to the following,

$$q_s^{\left[2\frac{L_{A_2}}{L}\right]} = F \cdot e^{-\alpha \bar{b}}, \quad (11)$$

$$\bar{D} = \frac{q_s^2}{L} \sum_{i \in A_2} \frac{W_i^2}{\beta_i} + \frac{1}{L} \sum_{i \in B_2} \sigma_i^2, \quad (12)$$

where

$$\begin{cases} \textit{Region } A_2: & \{\ i \in \Omega \text{ and } (\ \epsilon_i^2 \cdot \sigma_i^2) \geq (W_i^2 \cdot q_s^2/\beta_i\ )\ \} \\ \textit{Region } B_2: & \Omega\ -\ A_2\ , \end{cases}$$

where Ω = {0,1,...,L-1}, L_{A2} is the number of coefficients in Region A_2 and

$$F = \left[\prod_{i \in A_2} \left(\frac{\beta_i \cdot \epsilon_i^2 \cdot \sigma_i^2}{W_i^2}\right)\right]^{1/L}.$$

2.4. Threshold transform coder

In image coding we have to deal with not only the statistical behavior of the entire picture (objective criterion) but also the fidelity of individual samples embedded in their texture neighborhood (subjective criterion). Instead of selecting a fixed number of transform coefficients according to their average variances as suggested by the theory, we can also select the coded coefficients by their magnitudes, the so-called *threshold transform coding*. Assume that T_i is the threshold value used in picking up the i^{th} transform coefficient; that is, the i^{th} coefficient is set to zero before quantizing if its magnitude is less than T_i. In this case, both the α_i and the β_i parameters depend on the threshold value T_i. If we also include the bits to indicate the end of a block, b_{EOB}, then eqn (9) can be rewritten as follows.

$$\bar{b}(q_s) = \frac{1}{L}\left[\sum_{i=0}^{L-1} \frac{1}{\alpha_i(q_s, T_i)} \log_e\left(\frac{\gamma_i(q_s, T_i) \cdot \sigma_i^2}{W_i^2 \cdot q_s^2}\right) + b_{EOB}\right], \quad (13)$$

and

$$\bar{D} = \frac{q_s^2}{L} \sum_{i \in A_2} \left(\frac{W_i^2}{\beta_i(\triangle_i, T_i)}\right) + \frac{1}{L} \sum_{i \in B_2} \sigma_i^2. \quad (14)$$

If T_i is chosen to be $[constant \cdot W_i \cdot q_s]$ and about the same *constant* for all the coefficients then α, γ can be simplified to be functions of q_s. Replacing $\alpha_i(\cdot)$ in eqn (13) by $\alpha(\cdot)$, we obtain

$$q_s^2 = F(q_s) \cdot e^{-\alpha(q_s) \cdot \bar{b}}, \quad (15)$$

or

$$\bar{b}(q_s) = -\frac{1}{\alpha(q_s)} \log_e q_s^2 + \frac{G(q_s)}{\alpha(q_s)}, \quad (16)$$

where $F(q_s) = exp\{G(q_s)\}$.

We could further assume that the picture to be coded is not much different from the picture that has already been coded in the sense that α_i, β_i and γ_i remain about the same in the neighborhood of q_s that we are dealing with, then the F parameter in eqn (15) can be estimated from $\bar{b}$ and q_s of the coded pictures.

3. MODEL PARAMETERS

For a practical application, the parameters of the above model should be adjusted to cope with the underneath video coder structure and the real picture coding characteristics. The meaning of parameters in our model (eqns. (13) and (14)) suggests the following two modifications: First, the $\alpha(\simeq 1.386)$ is replaced by a parameter function $\alpha(q_s) = 1.386 \cdot \alpha_s(q_s)$ to include the non-ideal factors in a practical video coder. Second, the value of $\beta(\sigma/\triangle, T)$ as a function of $\sigma/\triangle$ is evaluated for low bit rate coding application.

3.1. Model Parameters β and α

The theoretical value of β is 12, which is approximately true when the quantization scale is much smaller than the signal variance. It can be shown that the $\beta(\sigma/\triangle, T)$ value for *Laplacian* probability density function (pdf) can be represented as $\frac{1}{\epsilon^2 \cdot (\frac{\sigma}{\triangle})^2} \cdot e^{\alpha \cdot H_Q[\frac{\sigma}{\triangle}, T]}$, where $H_Q[\frac{\sigma}{\triangle}, T]$ is the output entropy of uniform quantizer with quant. interval $\triangle$ and source variance σ^2.

It is rather complicated to compute $\beta(\frac{\sigma_i}{\triangle}, T)$ directly from the above equation. For simplicity we choose $T = \triangle$ and build a look-up table in the following simulations. Since the entropy $\simeq 0$ when $log_e(\frac{\triangle}{\sigma}) \geq 1.5$, the look-up table needs only to store β values for $log_e(\frac{\triangle}{\sigma})$ up to 1.5. For $log_e(\frac{\triangle}{\sigma}) \geq 1.5$, the corresponding distortion value is σ^2. In this case, from the original definition of β(eqn. (5)), β thus becomes $\frac{q_s^2 \cdot W_i^2}{\epsilon^2 \cdot \sigma_i^2}$. On the other hand, for the $\log_e(\triangle/\sigma) \leq -4$, a constant of 12 is assigned to β.

In Sec. 2, α is first viewed as a constant(=1.386), but the later analysis in the above suggests that α depends on the scale factors. We, for the sake of convenience, include the non-ideal factors in a practical coder through the use of the parameter function $\alpha(q_s)$. The values of these non-ideal factors for practical block transform coders are found from coding simulations on real pictures. We use the H.261-type coding structure

as an example and find that these non-ideal factors consists of *pdf* mismatch, non-ideal i.i.d assumption and inefficiency of the default VLC tables. Simulations show that if we let $\alpha(q_s)$ comprise these non-ideal factor as a whole, a first order linear equation is a good approximation of $\alpha(\cdot)$ for various q_s. This simplifies the calculation of our model.

3.2. Bits Prediction

So far, we obtain the parametric formulas of $\alpha(q_s)(= 1.386 \cdot \alpha_s(q_s))$ and $\beta(\sigma/\triangle)$. By combining these parametric formulas together with the model (eqns. (14), (15)), bits needed to encode a picture for a given distortion can be estimated if the variances of transform coefficients are available.

Since eqn (15) represents the relation between quantization scales and encoded bits, picture characteristics from the coding point of view can thus be represented by $F(q_s)$, which can be computed from variances of transform coefficients. We therefore call $F(q_s)$ *"coding complexity function"*.

4. VARIABLE FRAME RATE CODING

The $P \times 64k$ standard defines a Hypothetical Reference Decoder (HRD) model that all the standard compatible bit streams should comply with. In order to satisfy this HRD requirement, the Reference Model 8 (RM8) assumes a fixed frame rate, which is inefficient for low bit rate applications. At low bit rates, during periods of rapid motion, it is preferred to transmit fewer frames per second, but with a better quality for each transmitted frame. This is the basic operating principle behind VFR coding schemes.

The variable frame rate (VFR) coding problem addressed here can be described with the help of Fig. 2. The buffer/quantizer controller in Fig. 2 determines the new frame rate and the quantization scales to be used for the next coded frame, based on the information provided by the transform coefficients generator, entropy coder and output buffer. In a video sequence, except for scene changes, the same scene content usually lasts for several frames or it varies slowly. Since the picture to be coded is not much different from the picture that has just been coded, the $F(q_s)$ function in eqn. (15) should remain about the same value for the next frame and the value of $F(q_s)$ can be estimated from the $\bar{b}$ and q_s of the previously coded picture. Under this assumption, the controller can determine the number of skipped frames and the quantization scales for the next coded picture. In motion JPEG (MJPEG) coding, pictures are independently coded and thus can be skipped without propagation errors. We first apply the above source model to design a variable frame rate(VFR) scheme for MJPEG. Then a more complicated example of a VFR scheme for H.261 type intra/inter coding is also developed and simulated.

5. EXAMPLES AND SIMULATIONS

Four subsequences, *salesman, missa, claire* and *swing*, are concatenated into one as our test sequence to demonstrate the adaptation ability of our algorithm for different images and at scene changes.

5.1. Motion JPEG

In the simulations of variable frame rate MJPEG coding, results from using RM8 is also given for comparison, in which every picture is intra coded (denoted as RM8I). Fig. 3 shows the results using MJPEG coding on the test sequence at channel rate $P = 12$. It can be seen from Fig. 3 that a nearly constant level of quantization scales is provided by using the VFR scheme. Therefore, it produces a better image quality than the simple RM8I for approximately 3 to 6 dB PSNR improvement. The simulation results also show that the estimated bits using this model based on statistical data follows the picture variations(scene changes) quite well.

5.2. H.261-style intra/inter coding

The coding complexity functions for inter-coded pictures are difficult to estimate because their values depend on the quality of previously reconstructed pictures. However, the bits estimation model together with the VFR control algorithm can still track the variation of coding complexity functions that are difficult to compute directly. Simulation results in Fig. 4 shows that the coded image quality of VFR coding is controlled rather well and can be maintained at a higher level as compared to that of RM8. The VFR control is especially efficient at low bit rate coding, as can be seen from Table 1 that the PSNR improvement for channel rate P=2 is better then that of P=3 because at very low rate, precise bit rate control becomes very critical. More importantly, Table 1 shows that the PSNR with VFR coding is kept almost constant independent of picture contents variation.

6. CONCLUSION

We have derived a set of formulas that describe the relation between bits, distortion, and quantization stepsizes for transform coders. The realistic constraints

such as threshold coefficient selection are included in our formulation. Based on the proposed source model, we have developed a variable frame rate coding algorithm. With little computation overhead, this source model estimates coding bits quite accurately even at scene changes. Due to efficient bits estimation schemes, the coding control algorithm can reserve enough bits by skipping pictures to produce good coding quality. This coding control algorithm is quite simple and the simulation results confirm that an improved PSNR performance kept at a constant level can be achieved by using this VFR coding technique.

7. REFERENCES

[1] CCITT, Working Party XV/1, Draft of Recommendation H.261, Video Codec for Audiovisual Services at $P \times 64k$ bits/s, July 1990.

[2] William B. Pennebaker, JPEG - still image data compression standard. New York: VAN NOSTRAND REINHOLD, 1993

[3] MPEG (Motion Picture Experts Group) Video Committee Draft, ISO-IEC JTC1/SC2/WG11 Coding of Moving Pictures and Associated Audio, Dec 18, 1990.

[4] T. Berger, *Rate Distortion Theory* New Jersey: Prentice Hall, 1971.

[5] A.J. Viterbi and J.K. Omura, *Principles of Digital Communications and Coding*, New York: McGraw-Hill, 1979.

[6] H. Gish and J.N. Pierce, "Asymptotically Efficient Quantizing," *IEEE Trans. Inform. Theory*, pp.676 - 683, Sept 1968.

[7] N.S. Jayant and P. Noll, *Digital Coding of Waveforms*, New Jersey: Prentice Hall, 1984.

[8] A.N. Netravali and B.G. Haskell, *Digital Pictures: Representation and Compression*, New York: Plenum, 1988.

[9] Richard E. Blahut, *Principles and Practice of Information theory*. Addison Wesley, 1991.

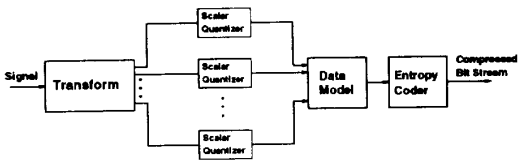


Figure 1: Practical image transform coder

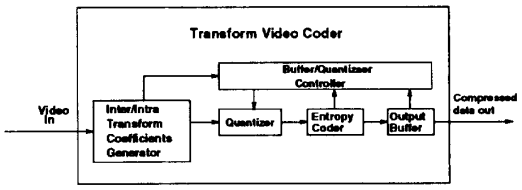
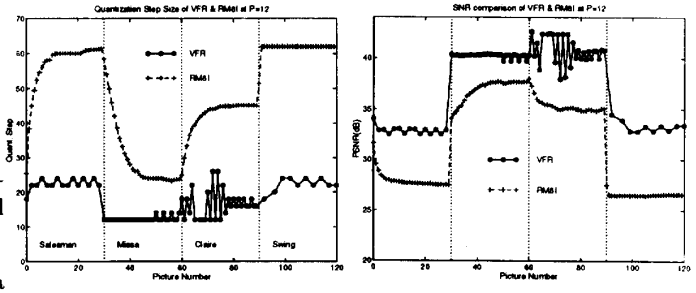


Figure 2: System diagram of variable frame coding



(a) (b)

Figure 3: Comparison of MJPEG VFR and RM8I in (a) quantization steps and (b) PSNR. Note that the number of "o" (VFR) are less than that of "+"(RM8I) since some pictures are skipped in the VFR coding

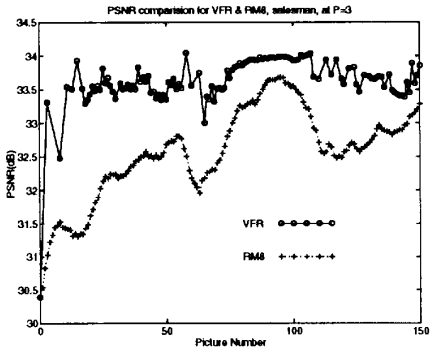


Figure 4: PSNR comparison of H.261-type intra/inter VFR coding over RM8 on "salesman" at P=3.

VFR(RM8)

PSNR(dB)		Salesman	Missa	Claire	Swing
Average	P=2	33.89(30.80)	39.09(38.56)	40.36(38.83)	34.83(31.97)
	P=3	34.01(31.55)	39.40(39.15)	41.16(40.24)	35.23(32.89)
Variance	P=2	0.42(0.06)	0.05(0.42)	0.12(0.45)	0.21(0.73)
	P=3	0.15(0.18)	0.04(0.53)	0.16(1.05)	0.27(1.48)
Peak difference	P=2	1.99(0.80)	0.83(2.73)	1.26(2.59)	1.77(2.88)
	P=3	1.84(1.87)	0.73(3.23)	1.72(4.29)	2.39(4.20)

Table 1: Comparisons of VFR and RM8 at P=2 and P=3 for average, variance and peak difference of PSNR.