Practical Issues of Dynamic Bit Allocation in Multimedia Source Compression

Kai-Kuang Ma and Dinesh P. Mital

Nanyang Technological University

School of Electrical and Electronic Engineering, Nanyang Technological University, Nanyang Avenue, Singapore 639798, Republic of Singapore.

Telephone: (65) 799-6366. Fax: (65) 792-0415.

email: ekkma@ntuvax.ntu.ac.sg and emital@ntuvax.ntu.ac.sg

Invited Paper

Abstract

Owing to various emerging multimedia applications which create enormous amount of data traffic over limited channel bandwidth, dynamic bit allocation plays a significant role on achieving robust and/or optimum source coding performance. In this paper, we focus on *pixel-based* and *perceptual-based* dynamic bit allocations which have been widely exploited in many source coding algorithms and international audio/visual compression standards as well. Many fundamental facts of human auditory system have been effectively exploited in perceptual-based bit allocation for speech and audio signals. On the contrary, pixel-based bit allocation which is widely used in image and video coding has certain practical issues required to be addressed. We highlight these observed concerns, present an unified and optimum solution, and recommend a future direction.

Keywords

Dynamic bit allocation, multimedia, image coding, audio coding, subband coding, perceptual coding, psychoacoustical model, MPEG.

1 INTRODUCTION – SOURCE CODING AND BIT ALLOCATION

Source coding or compression is one of key elements which has been significantly contributing today's multimedia technology and its future outgrowth as well. Among var-

Figure 1 A general subband source coding system

ious powerful source coding techniques, discrete cosine transform (DCT) based coding (Rao and Yip 1990), subband/wavelet coding (Woods and O'Neil 1986, Antonini *et al.* 1992, Vaidyanathan 1993, Vetteli and Kovačević 1995), and/or psychophysical- and psychoacoustical-model based perceptual coding (Pan 1995, ISO/IEC 1993, Jayant *et al.* 1993) have been widely viewed as the mainstream of compression technology and exploited in multiple international compression standards such as JPEG, H.261, H.263, MPEG/Video, and MPEG/Audio for various speech, image, video, and audio applications (Bhaskaran and Konstantinides 1996, Netravali and Haskell 1995).

Dynamic bit allocation is one of the major subsystems of aforementioned coding systems which distributes a given quota of bits to the concerned sources. In the context of DCT-based or subband/wavelet coding, these 'sources' are the DCT coefficients or critically down-sampled subbands respectively. To perceptual coding (*e.g.*, MPEG-1/Audio coding for audio signals), thirty-two subbands are first generated using polyphase subband decomposition scheme or modified discrete cosine transform (MDCT) (Pan 1995, ISO/IEC 1993, Bhaskaran and Konstantinides 1996). A general multiband source coding framework is presented in Figure 1.

Instead of processing one full-band signal (*i.e.*, the original source), multiple smaller subbands as described above, jointly provides a more flexible and adaptive structure for more efficient dynamic bit allocation. As a result, better coding performance could be achieved at the expense of increased complexity and computations.

2 PIXEL-BASED DYNAMIC BIT ALLOCATION

In general, dynamic bit allocation can be classified as *pixel-based* or *perceptual-based*. We shall discuss the former first.

Figure 2 Subband tiling: (a) Equal bandwidth, (b) Octave bandwidth, (c) Arbitrary bandwidth

2.1 Direct Form of Arbitrary Subband Tiling

Assume that K subbands were decomposed from the input source using quadrature mirror filter (QMF) bank (Vaidyanathan 1993, Vetteli and Kovačević 1995). There are various ways to partition the full-band source. Equal-bandwidth tiling as shown in Figure 2(a) has been suggested as the best approach for image coding; typically, 16 subbands for a 256×256 or 512×512 image. Octave bandwidth partition is realized by recursively decompose the resulting baseband and used in wavelet coding. Arbitrary bandwidth tiling is newly recommended for adaptive subband image coding.

By imposing Shannon's rate-distortion bounds (Shannon 1963) to all K subbands and applying Lagrange technique, the general solution of optimum bit allocation for any subband j with bandwidth w_j and average power $\mathcal{P}_j$ is derived as:

$$b_j^* = \left(\frac{w_j}{\mathcal{W}}\right) \mathcal{B} + \sum_{i=1}^K \left(\frac{w_i w_j}{\mathcal{W}}\right) \log_2 \left(\frac{\mathcal{P}_j w_i}{\mathcal{P}_i w_j}\right), \quad \text{for } j = 1, 2, \dots, K; \quad (1)$$

where $\mathcal{B}$ and $\mathcal{W}$ are the given total bit rate and the total bandwidth, respectively.

For full derivation details, refer to (Ma 1992). Equation (1) is called the *direct form* bit allocation which solely depends on pixel intensity. For special cases such as equal bandwidth, octave bandwidth, and weighting-factor incorporated bit allocation can be found in (Ma and Rajala 1996, 1992), (Ma 1994b), and (Ma and Rajala 1994a), respectively.

2.2 Practical Issues

Some practical concerns are inevitably incurred in pixel-based bit allocation as described in the following.

Equation (1) could cause negative bit allocation; thus, not realizable.

Second, a mismatch often occurs between calculated results using Equation (1) and the quantizer's bit rate. The difference between the two is the *unassigned bit rate*. Their sum from all the subbands, called *total unassigned bit rate*, is equal to $\mathcal{B} - \sum_j b_j^*$ and is usually quite large which causes unacceptable imaging quality.

Third, for different set of subband quantizers, the entire derivations has to start all over again. This would significantly increase the difficulty of adaptively changing the subband quantizer set in real-time applications.

Fourth, adaptiveness could be exploited into source model as well. In such case, arbitrary subband tiling or partition based on some partition criteria could be implemented into the source for adaptive subband decomposition.

Fifth, signal parameters extracted from the subbands are also greatly affecting bit allocation result. Refer to a simulation demonstration presented in Section 3.2.

3 SOLUTION AND SIMULATION

3.1 Sequential Form – An Unified and Optimum Solution

Heuristic or ad-hoc approaches have been previously attempted for solving the first two issues addressed in Section 2.2 while the solution of the remaining issues was unknown until a *sequential form* algorithm was developed by (Ma and Rajala 1996, 1992) which solves all above-mentioned practical concerns and is generalized and optimized to use.

Ma's sequential form is an iterative implementation algorithm of the *direct form*. The sequential processes were derived in such a way that optimum bit assignment is maintained in each iteration. For the algorithmic details, refer to the flowchart documented in (Ma and Rajala 1992, 1996).

3.2 Bit Allocation Performance in Wavelet Image Coding

To demonstrate the bit allocation performance of using sequential form in wavelet image coding, a set of absolute moment block truncation coder (AMBTC), developed by (Lema and Mitchell 1984), are chosen as the subband quantizers. In such framework, the objective of dynamic bit allocation algorithm equivalently decides an AMBTC window size for each subband. The corresponding assigned AMBTC windows are listed in Table 1.

Note that Ma's sequential form uses subband *mean energy* as the parameter, but (Huang and Schultheiss 1963) exploits subband *variance*. Using identical simulation framework documented in (Ma 1992–6), two images are produced and shown in Figure 3.

Despite that there is only a small peak signal-to-noise ratio (PSNR) difference (0.2 dB)

Subband Label:	1	2	3	4	5	6	7	Unassigned Bits
(a) Variance-Based	1	2	0	4	8	0	0	0.0
(b) Energy-Based	1	1	4	4	0	0	0	0.0

Table 1 Subband bit allocation results in terms of the AMBTC window assignment for the images presented in Figure 3. For example, table entry "8" in subband 5 of (a) means an 8×8 AMBTC window was applied to that band. Entries "1" and "0" mean that the corresponding subbands will be *uncompressed* and *discarded*, respectively. For subband labeling, refer to Figure 2(b).

Figure 3 (a) | (b) Two reconstructed LENA images at 1.25 *bpp* using: (a) Huang's *variance*-based method — 29.14 dB and (b) Ma's *energy*-based method — 29.34 dB. The original (uncompressed) LENA image has a size of 256×256 with 8 *bpp*.

between Figure 3(a) and Figure 3(b), using subband mean energy as the parameter for bit allocation has clearly achieved superior imaging quality.

4 PERCEPTUAL-BASED DYNAMIC BIT ALLOCATION

Since the human being is the final judge, certain well-known properties of the human visual system (HVS) and human auditory system (HAS), such as masking effect, would

Figure 4 MPEG-1/Audio Codec System: (a) Encoder and (b) Decoder.

be highly useful to be incorporated into the design of efficient bit allocation algorithms. Such perceptual-based bit allocation has been adopted in the MPEG-1/Audio standards (Pan 1995, ISO/IEC 1993) where two psychoacoustical models, Model I and Model II, are suggested for achieving *perceptually transparent* CD-quality audio and/or wideband speech at compression ratios around 6 to 12.

Refer to Figure 4, the psychoacoustical Model I computes the fast Fourier transform (FFT) for consecutive segments of input audio samples followed by a nonlinear mapping into the critical bands (with scale in "bark"). The masking effect of HAS is then exploited. Refer to Figure 5, the *threshold-in-quiet* curve (in dotted line) is the absolute threshold for the human hearing below which any tone becomes inaudible.

At the presence of a strong tonal signal ("*masker*") with sound pressure level (SPL) exceeded the threshold-in-quiet curve boundary causes a sharp raise of the threshold curve around its neighborhood. As a result, the adjacent tones ("*maskees*") which were audible before masking are possible to be inaudible after masking, depending on whether each maskee's SPL exceeds the new threshold-in-quiet (in solid line). For example, inaudible tones 'D' and 'G' in Figure 5 were audible before masking.

After calculating signal-to-mask ratio (SMR) by the Model I (involving 9 steps), the

Figure 5 Perceptual-based dynamic bit allocation. The allocated bits for tones 'C', 'F', and 'H' were computed according to audibility, in terms of signal-to-masking ratio (SMR), of subband signals.

bits are then allocated based on their 'effective' tonal strength (ISO/IEC 1993); i.e., the overshoot above the masking threshold.

5 CONCLUSION

In this paper, we provide an overview of pixel-based and perceptual-based dynamic bit allocations for digital source coding. Observed practical issues are highlighted and solved by the sequential form developed by (Ma and Rajala 1992–6) which performs optimum bit allocation and achieves attractive coding performance.

Although we only use still images for demonstration here, the presented bit allocation algorithm can be directly exploited into other coding framework (*e.g.*, DCT-based) and digital media.

On the emerging wireless communication applications, a low or very low bit rate is usually desired. Therefore, a sophisticated dynamic bit allocation will even have a more significant impact on the coding performance at such low bit-rate applications. For that, we suggest here that the hybrid approach, which combines the pixel-based and perceptual-based bit allocations, should be further explored; *e.g.*, cascading them in serial.

REFERENCES

- Rao, K.R. and Yip, P. (1990) *Discrete Cosine Transform - Algorithms, Advantages, Applications*. San Diego, CA: Academic Press.
- Woods, J.W. and S.D. O'Neil, S.D. (1986) Sub-Band Coding of Images. *IEEE Trans. Acoustics, Speech, and Signal Proc.*, **ASSP-34**, 1278-88.
- Antonini, M., Barlaud, M., Mathieu, P., and Daubechies, I. (1992) Image Coding Using Wavelet Transform. *IEEE Trans. on Image Proc.*, **Vol. 1, No. 2**, 205-20.
- Vaidyanathan, P.P. (1993) *Multirate Systems and Filter Banks*. Englewood Cliffs, NJ: Prentice-Hall.
- Vetterli, M. and Kovačević, J. (1995) *Wavelets and Subband Coding*. Englewood Cliffs, NJ: Prentice-Hall.
- Pan, D. (1995) A Tutorial on MPEG/Audio Compression. *IEEE Multimedia*, **Vol. 2, No. 2**, 60-74.
- ISO/IEC (1993) *Information Technology — Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbits/s — IS 11172-3 (Part3, Audio)*
- Jayant, N., Johnston J., and Safranek, R. (1993) Signal Compression Based on Models of Human Perception. *Proceedings of the IEEE*, **Vol.81, No. 10**, 1385-422.
- Bhaskaran, V. and Konstantinides, K. (1996) *Image and Video Compression Standards - Algorithms and Architectures*. Boston, MA: Kluwer Academic Publishers.
- Netravali, A.N. and Haskell, B.G. (1995) *Digital Pictures: Representation, Compression, and Standards*. Second Edition. New York, NY: Plenum Press.
- Shannon, C.E. and Weaver, W. (1963) *The Mathematical Theory of Communication*. Urbana, IL: University of Illinois Press.
- Ma, K.-K. (1992) *Subband Image Coding Using Absolute Moment Block Truncation with A Shannon-Bound Sequential Dynamic Bit Allocation Algorithm*, Ph.D. dissertation, North Carolina State University, Raleigh, North Carolina, USA, 1992.
- Ma, K.-K. and Rajala, S.A. (1996) Subband Absolute Moment Block Truncation Coding. *Optical Engineering* — Special Issue Visual Communications and Image Processing, **Vol. 35, No. 1**, 213-31.
- Ma, K.-K. (1994a) Generalized Optimum Dynamic Bit Allocation Scheme for Source Compression. *IEEE International Conf. on Image Processing*, 864-8.
- Ma, K.-K. (1994b) New Bit Allocation Method for Wavelet and Octave Subband Coding. *IEEE Asia-Pacific Conf. on Circuits and Systems*, 624-9.
- Huang, J.J.Y. and Schultheiss, P.M. (1963) Block Quantization of Correlated Gaussian Random Variables. *IEEE Trans. on Commun.*, **Vol. CS-11**, 289-96.
- Lema, M.D. and Mitchell, O.R. (1984) Absolute Moment Block Truncation Coding and Its Application to Color Images. *IEEE Trans. on Commun.*, **Vol. 32, No. 10**, 1148-57.