Submitted to ICIP'97

LEXICOGRAPHIC BIT ALLOCATION FOR MPEG VIDEO CODING

Dzung T. Hoang[*]
Digital Video Systems, Inc.
2710 Walsh Ave., Ste. 200
Santa Clara, CA 95051
dth@dvsystems.com

Jeffrey Scott Vitter[†]
Duke University
Box 90129, Computer Science
Durham, NC 27708–0129
jsv@cs.duke.edu

Elliot L. Linzer[‡]
C-Cube Microsystems
One Water Street, 2nd Floor
White Plains, NY 10601
elliot.linzer@c-cube.com

ABSTRACT

We consider the problem of allocating bits among pictures in an MPEG video coder to equalize the visual quality of the coded pictures, while meeting buffer and channel constraints imposed by the MPEG Video Buffering Verifier. We address this problem within a framework that consists of three components: 1) a bit production model for the input pictures, 2) a set of bit-rate constraints imposed by the Video Buffering Verifier, and 3) a novel lexicographic criterion for optimality. Under this framework, we derive simple necessary and sufficient conditions for optimality that lead to efficient algorithms.

1. INTRODUCTION

In addition to specifying a syntax for the encoded bitstream and a mechanism for decoding it, the MPEG standards define an idealized decoder, called the Video Buffering Verifier (VBV), that puts quantifiable limits on the variability in the coding rate. In a typical MPEG encoder, VBV compliance is handled by the rate controller, whose purpose is to allocate bits to coding units and to regulate the coding rate to meet the constraints imposed by the VBV while maintaining an acceptable level of quality. In this paper, we consider rate control in the context of the MPEG-1 and MPEG-2 standards for video coding.

Previous work in optimal buffer-constrained rate control generally seeks to minimize a sum-distortion measure, typically mean-squared error (MSE), averaged over coding blocks [1]. While this approach leverages a wealth of tools from optimization theory and operations research, it does not guarantee the constancy in quality that is generally desired from a video coding system. For example, a video sequence with a constant or near-constant level of distortion is more desirable than one with lower average distortion but higher variability. This is because human viewers tend to find frequent changes in quality more noticeable and annoying.

We propose a novel optimality criterion that better expresses the desired constancy of quality. The idea is to minimize the maximum block distortion and then minimize the second highest block distortion, and so on. The intuition is that doing so would equalize distortion by limiting peaks in distortion to their minimum. This criterion is referred to as *lexicographic optimality* or *lexicographic minimax* in the literature [2].

2. LEXICOGRAPHIC FRAMEWORK

In this paper, we introduce a framework for bit allocation for video coding under VBV constraints and with a total bit budget. The framework consists of three components: 1) a bit-production model, 2) a set of buffer constraints for constant and variable bit rate operation, and 3) a novel lexicographic optimality criterion. We formalize bit allocation as a resource allocation problem with continuous variables and non-linear constraints, to which we apply a global lexicographic optimality criterion.

Analysis of the framework for constant and variable bit rate operation reveals a set of simple and elegant conditions for optimality that admit efficient algorithms. In this paper, we provide a summary of and intuitions behind the analysis. For details of the analysis and algorithms, the reader is referred to [3].

[*] Work was begun when the author was at IBM T. J. Watson Research Center.

[†] Support was provided in part by Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grants F49620–92–J–0515 and F49620–94–1–0217, by Army Research Office grant DAAH04–93–G–0076, and by an associate membership in CESDIS.

[‡] Work was performed when author was at IBM T. J. Watson Research Center.

2.1. Perceptual Quantization

In a typical video coder, the output bit rate can be regulated by adjusting a *quantization scale* Q_s. Coding with a constant value of Q_s generally does not result in either constant bit rate or constant perceived quality. Studies into human visual perception suggest that perceptual distortion is correlated to certain spatial (and temporal) properties of an image (video sequence). These studies lead to various quantization techniques, called *perceptual quantization* or *adaptive quantization*, that take into account properties of the Human Visual System [4, 5, 6].

Based upon this body of work, we propose a separation of the quantization scale Q_s into a *nominal quantization* Q and a *perceptual quantization function* $P(I, Q)$ such that $Q_\mathrm{s} = P(I, Q)$, where I denotes the block being quantized. The function P is chosen so that if the same nominal quantization Q were used to code two blocks then the blocks would have the same perceptual distortion. In this way, the nominal quantization parameter Q would correspond directly to the perceived distortion and can serve as the object for optimization. We favor a multiplicative model where $P(I, Q) = \alpha_\mathrm{I} Q$, such that α_I is large where quantization noise is less noticeable. Our bit rate allocation, however, works with any perceptual quantization function.

2.2. Bit-Production Modeling

For a sequence of N pictures, we define N independent bit-production models $\{f_1, f_2, \ldots, f_N\}$ that map nominal quantization scale to bits: $b_i = f_i(q_i)$, where $f_i : [0, \infty] \mapsto [l_i, u_i]$, with $0 \leq l_i < u_i$.[1] We restrict the models to be continuous and monotonically decreasing. Although there are cases where monotonicity does not hold in practice, it is a generally accepted assumption.

The number of bits produced for a picture also depends upon a myriad of coding choices besides quantization scale, such as motion compensation and the mode used to code each block. We assume that these choices are made independently of quantization and prior to performing rate control.

2.3. VBV Buffer Constraints

The MPEG standards define a hypothetical decoder, called the Video Buffering Verifier (VBV), that consists of a decoder buffer, a decoder, and a display unit. The decoder buffer stores the incoming bits for processing by the decoder. At regular display intervals, the decoder *instantaneously* removes, decodes, and displays the earliest picture in the buffer.

The VBV has two prescribed modes of operation: *constant bit rate* (CBR) and *variable bit rate* (VBR). MPEG-1 supports only CBR mode while MPEG-2 supports both modes. In CBR mode, bits enter the decoder buffer at a constant rate R_max, specified in the compressed stream. Initially, the buffer is empty and fills for a prespecified time before bits for the first picture are removed and decoded. Afterwards, the buffer continues to fill at the channel rate R_max while the decoder removes bits for coded pictures at regular display intervals. The CBR mode models operation of a decoder connected to a constant-bit-rate channel with a fixed channel delay.

Denoting the buffer fullness just before picture i is removed from the VBV buffer using allocation s by $B_\mathrm{f}(s, i)$, we can describe CBR operation of the VBV buffer by the following recurrence.

$$\begin{aligned} B_\mathrm{f}(s, 1) &= B_1 \\ B_\mathrm{f}(s, i+1) &= B_\mathrm{f}(s, i) + B_\mathrm{a}(i) - s_i \end{aligned}$$

Here, B_1 is the initial buffer fullness, $B_\mathrm{a}(i)$ the number of bits that enter the buffer after picture i is removed, and s_i the number of bits allocated to picture i.

In VBR mode, the decoder buffer is initially filled to capacity at the peak rate R_max before the first picture is removed. Thereafter, bits enter the buffer at the peak rate until the buffer is full. When the buffer is full, bits stop entering the buffer until the next picture has been decoded. Since the decoder buffer stops receiving bits when it is full, a potentially variable number of bits can enter the buffer in each display period. The VBR mode can be thought of as modeling the operation of a decoder connected to a channel or device, a disk drive for example, that can transfer data at a variable rate up to the peak rate R_max.

VBR operation of the VBV buffer can be described with the following recurrence.

$$\begin{aligned} B_\mathrm{f}(s, 1) &= B_1 \\ B_\mathrm{f}(s, i+1) &= \min\{B_\mathrm{vbv}, B_\mathrm{f}(s, i) + B_\mathrm{a}(i) - s_i\} \end{aligned}$$

The effect of the minimization is to prevent the buffer from overflowing.

For proper operation in either CBR or VBR mode, the decoder buffer should not exceed its capacity B_vbv.[2] Also, the buffer should contain at least the number of bits needed to decode the next picture at the time it is to be decoded; i.e., $B_\mathrm{f}(s, i) \geq s_i$. These requirements impose upper and lower bounds on the number of bits that the encoder can produce for each picture.

[1] We number pictures in encoding order and not in display order.

[2] By definition, this requirement is always met in VBR mode.

2.4. Lexicographic Optimality

We now formally define the lexicographic optimality criterion. Let S be the set of all legal allocations for a bit-allocation problem P. For a bit allocation $s = \langle s_1, s_2, \ldots, s_n \rangle \in S$, let $\mathbf{Q}^s = \langle Q_1, Q_2, \ldots, Q_N \rangle$ be the sequence of Q-values to achieve the bit allocation specified by s; i.e., $Q_i = f_i^{-1}(s_i)$.

We define a permutation DEC on $\mathbf{Q}^s$ such that for $\mathrm{DEC}(\mathbf{Q}^s) = \langle q_{j_1}, q_{j_2}, \ldots, q_{j_N} \rangle$, we have $q_{j_1} \geq q_{j_2} \geq \cdots \geq q_{j_N}$. We define $\mathrm{rank}(s,k)$ to be the k^{th} element of $\mathrm{DEC}(\mathbf{Q}^s)$; i.e., $\mathrm{rank}(s,k) = q_{j_k}$. We define a binary relation $\prec$ on allocations as follows: $s \prec s'$ if and only if $\mathrm{rank}(s,j) = \mathrm{rank}(s',j)$ for $j = 1, 2, \ldots, k-1$ and $\mathrm{rank}(s,k) < \mathrm{rank}(s',k)$ for some $1 \leq k \leq N$. We define $s \asymp s'$ if and only if $\mathrm{rank}(s,j) = \mathrm{rank}(s',j)$ for all j. Similarly we define $s \preceq s'$ if and only if $s \prec s'$ or $s \asymp s'$.

Definition 1 A legal allocation s^* is *lexicographically optimal* if $s^* \preceq s$ for all other legal allocation s.

Lemma 1 (Constant-Q) *Given a bit-allocation problem P of length N, if there exists a legal allocation $s = \langle s_1, s_2, \ldots, s_N \rangle$ such that $f_n^{-1}(s_n) = q$ for all n, then s is the only lexicographically optimal allocation for P.*

This lemma establishes a desirable property of lexicographic optimality: If a constant-Q allocation is legal, it is the only lexicographically optimal allocation. This meets our objective of obtaining a constant-quality allocation when feasible.

3. CBR ANALYSIS

Analysis of the lexicographic framework under CBR constraints yields a set of necessary and sufficient conditions, as summarized in the following theorem.

Theorem 1 *Given a CBR bit-allocation problem P of length N, a legal allocation $s = \langle s_1, s_2, \ldots, s_N \rangle$ is optimal if and only if the following conditions hold. Also, the optimal allocation is unique.*

1. *If $Q_j > Q_{j+1}$ for $1 \leq j < N$, then $B_{\mathrm{f}}(s,j) = s_j$.*
2. *If $Q_j < Q_{j+1}$ for $1 \leq j < N$, then $B_{\mathrm{f}}(s,j+1) = B_{\mathrm{vbv}}$.*

In an optimal allocation, if Q changes from one picture to the next, the VBV buffer must be in one of two states: *empty* or *full*. If Q increases from picture j to picture $j+1$, the buffer must be be full immediately before picture $j+1$ is removed from the buffer. If Q decreases, the buffer must be empty immediately after picture j is removed. These results confirm the intuition that, to reduce any difference in perceptual distortion between two consecutive pictures, one needs to be able to "shift" bits from the easier-to-code picture to the harder-to-code picture. The extent that bits can be shifted is limited by the VBV buffer.

Theorem 1 guarantees that an allocation that meets the specified "switching" conditions and does not cause the buffer to underflow or overflow is lexicographically optimal. Using these results, we can design a dynamic programming algorithm to compute an optimal allocation. The basic idea behind dynamic programming is to decompose a given problem in terms of optimal solutions to smaller problems. All we need to do is maintain invariant the conditions stated in Theorem 1 for each subproblem we solve. We do this by incrementally constructing optimal bit allocations for pictures 1 to k that results in the VBV buffer being either *full* or *empty* after picture k is decoded. These states are exactly the states where a change in Q may occur. Once we have computed the above optimal allocations for pictures 1 to m, where $m \leq k$, we can compute an optimal allocation for pictures 1 to $k+1$ with a desired endpoint (full or empty) by searching for a constant-Q segment that joins the endpoint of a precomputed allocation and the desired endpoint for picture $k+1$.

With the above algorithm, $O(N)$ space is used and $O(N^2)$ possible allocations are considered. For most practical bit-production models, computing a constant-Q segment requires time linear in the length of the segment, yielding a total running time of $O(N^3)$. For some useful classes of bit-production models, such as the hyperbolic $f_i(q_i) = \alpha_i/q_i + \beta_i$, we can compute a constant-Q segment in constant amortized time, resulting $O(N^2)$ time complexity.

4. VBR ANALYSIS

The total number of bits that a CBR encoding can use is dictated by the channel bit rate and the buffer size. With VBR operation, the total number of bits has no lower bound, and its upper bound is determined by the peak bit rate and the buffer size. Consequently, VBR is useful and most advantageous over CBR when the average bit rate needs to be lower than the peak bit rate. This is especially critical in storage applications, where the storage capacity, and not the transfer rate, is the limiting factor.

For typical VBR applications bits enter the decoder buffer at an effective rate that is less than the peak during the display interval of many pictures. In interesting cases, there will be segments of pictures that are best coded with an average bit rate that is higher than the peak. This is possible because of the buffering. During the display of these pictures, bits enter the VBV buffer at the peak rate. Since these pictures consume bits at a rate higher than the average, they are conceptually "harder" to code than the other "easier" pictures.

To equalize quality, easy pictures should be coded at the same base quality. It does not pay to code a hard picture at a quality higher than that of the easy pictures; the bits expended to do so could be better distributed to raise the quality of the easy pictures. Therefore the base quality for the easy pictures should also serve as the maximum quality level for the sequence. Hard pictures, on the other hand, are limited by the peak rate and are effectively operating under CBR constraints, with the CBR rate being the peak VBR rate. Consequently, hard pictures should be coded as if in CBR mode.

Analysis of the lexicographic framework under VBR constraints validates the above intuitions and is summarized in the following theorem.

Theorem 2 *Given a VBR bit-allocation problem P of length N, a legal allocation $s = \langle s_1, s_2, \ldots, s_N \rangle$ is optimal if and only if the following conditions hold. Also, the optimal allocation is unique.*

1. *If $B_{\mathrm{f}}(s,j) + B_{\mathrm{a}}(j) - s_j > B_{\mathrm{vbv}}$ for $1 \le j \le N$, then $Q_j = \min_{1\le k\le N}\{Q_k\}$.*
2. *If $B_{\mathrm{f}}(s^*,N) > s^*_N$ then $Q_N = \min_{1\le k\le N}\{Q_k\}$.*
3. *If $Q_j > Q_{j+1}$ for $1 \le j < N$, then $B_{\mathrm{f}}(s,j) = s_j$.*
4. *If $Q_j < Q_{j+1}$ for $1 \le j < N$, then $B_{\mathrm{f}}(s,j+1) = B_{\mathrm{vbv}}$ and $B_{\mathrm{f}}(s,j+1) + B_{\mathrm{a}}(j+1) - s_{j+1} \le B_{\mathrm{vbv}}$.*

Computing an optimal VBR allocation now reduces to identifying easy and hard pictures and finding the minimum quantization scale needed to meet the target bit budget. Further analysis in [3] shows that this can be done within a finite number of iterations of a simple algorithm that simulates the operation of the VBV to determine hard and easy pictures and invokes the CBR algorithm for hard pictures. Analysis of the VBR algorithm yields the same time and space complexity as the CBR algorithm.

5. EXPERIMENTAL RESULTS

We implemented the CBR and VBR bit allocation algorithms within a software MPEG-2 encoder. A linear spline is used to model the bit production. Multiple encoding passes, each using a fixed nominal quantization, are used to construct the model. In the final encoding pass, the optimal bit allocation algorithms are used to compute the nominal quantization for each picture. To recover from errors in the model, the bit allocation is recomputed after each picture is coded.

To test the effectiveness of lexicographic bit allocation, we performed coding simulations on a two-minute video clip that contains scenes of widely varying complexities, from a static block diagram to a high-action sports scene. The video sequence is in NTSC CCIR-601 format. We code the sequence in CBR mode at 3 Mbit/s and in VBR mode at 3 Mbit/s average and 6 Mbit/s peak. The VBV buffer size is 1,835,008 bits. The results are summarized in Table 1. Results for optimal CBR allocation show less variance in PSNR and considerably less in nominal quantization than results obtained using the MPEG Test Model 5 [7]. The results for optimal VBR allocation show nearly constant nominal quantization and much less variance in PSNR.

	PSNR				Nominal Q	
	Avg	Var	Min	Max	Avg	Var
TM5	33.5	31.7	21.4	66.1	16.3	232.3
CBR	33.6	30.3	22.8	62.7	13.9	96.3
VBR	32.4	4.2	27.4	62.8	13.7	0.076

Table 1: Summary of encoding simulations.

6. REFERENCES

[1] A. Ortega, K. Ramchandran, and M. Vetterli. Optimal trellis-based buffered compression and fast approximations. *IEEE Trans. on Image Processing*, 3(1):26–40, January 1994.

[2] T. Ibaraki and N. Katoh. *Resource Allocation Problems*. MIT Press, Cambridge, MA, 1988.

[3] D. T. Hoang, E. Linzer, and J. S. Vitter. Lexicographically optimal rate control for video coding with MPEG buffer constraints. Technical Report CS-1996-02, Duke University, Dept. of Computer Science, 1996.

[4] E. Viscito and C. Gonzales. A video compression algorithm with adaptive bit allocation and quantization. In *SPIE Proceedings: Visual Communications and Image Processing*, volume 1605, pages 58–72, November 1991.

[5] N. S. Jayant, J. Johnson, and R. Safranek. Signal compression based on models of human perception. *Proc. of the IEEE*, 81:1385–1422, October 1993.

[6] T.-Y. Chung, K.-H. Jung, Y.-N. Oh, and D.-H. Shin. Quantization control for improvement of image quality compatible with MPEG2. *IEEE Trans. on Consumer Electronics*, 40(4):821–825, November 1994.

[7] ISO-IEC/JTC1/SC29/WG11/N0400. Test model 5, April 1993. Document AVC-491b, Version 2.