Trellis-Based R-D Optimal Quantization in H.263+

Max Luttrell, Jiangtao Wen
PacketVideo Corporation
luttrell, gwen@packetvideo.com

John D. Villasenor
Electrical Engineering Department
University of California, Los Angeles
villa@icsl.ucla.edu

May 23, 2000

Abstract:

We describe a trellis-based algorithm which enables R-D optimum quantization decisions in the H.263+ [1] video coding standard. The use of the trellis allows the quantization decisions for all coefficients in a block to be made jointly, and constrasts with more commonly used R-D optimizations which operate on one coefficient at a time. Experiments conducted using H.263+ for video coding rates of 40-50 kbps show an average improvement of 3.5% in bitrate, or equivalently 0.17dB in PSNR, relative to implementations which follow the International Telecommunications Union (ITU) Test Model specifications in making quantization decisions.

1 Introduction

Video coding algorithms including the ITU H.26x family of video coding standards and the ISO MPEG standards all utilize the discrete cosine transform (DCT), scalar quantization of DCT coefficients, and run-length and Huffman coding of the quantized coefficients. This framework offers significant opportunities to apply rate-distortion considerations at the encoder. In this paper, we describe a rate-distortion optimal approach to quantization decisions in H.263+, more formally known as H.263 version 2. The algorithm we present is based on dynamic programming, and like other algorithms in this class such as the Viterbi algorithm, uses a trellis to identify a path which minimizes a metric which in this case is the cost function $R+\lambda D$ where R is rate and D is distortion.

The basic idea underlying rate-distortion optimal coding, as described in many previous publications including [2] and [3], is to define and then minimize a cost function $R + \lambda D$ such that both the rate R and distortion D are considered in coding decisions. λ is a user-determined parameter which selects the operating location on the rate-distortion curve. In limiting case of large λ, only the distortion is considered, and as λ approaches 0, coding decisions are made with rate minimization as the only criterion.

Optimizations of this type have a long history in coding theory. In the video coding standards and other algorithms involving run-length coding, however, the rate needed to describe a particular coefficient c_k is a function not only of the quantization decision made for c_k but also of the decisions made on the neighboring coefficients $c_i, i < k$ and $c_i, i > k$. For example, the rate needed to express a coefficient that is quantized to zero depends on the length of the "run" which will be created by such a quantization decision. It is still possible to make R-D optimal decisions in systems including run-length coding, though the publications that have addressed this have dealt mostly with systems that differ quite significantly from H.263+.

The dependencies between adjacent coefficients in DCT-based video coding means that the number of possible joint quantization decisions and corresponding (R, D) pairs grows exponentially with the number of coefficients under consideration. Dynamic programming techniques offer a means to find the optimum set of decisions in linear time when the goal, as in this case, is minimization of $R + \lambda D$ for the whole block.

0-7803-6297-7/00/$10.00

We describe the construction and use of a trellis which identifies the optimum set of quantization decisions for each block and provide a set of experimental results using H.263+. We focus on bitrates in the range of 40-50 kbps as this is high enough to enable sufficient quality for videoconferencing and other applications, while also low enough for low-cost transmission over a wide range of existing and emerging wireline and wireless systems.

2 Trellis Construction and Processing

The algorithm presented focuses on producing the optimal set of quantization decisions for each block. Descriptions of video coding and in particular of the algorithms such as MPEG-2, H.263, and H.263+ are found in references including [4] and [5], and are not included here due to space constraints. We focus our attention on H.263+, as this is the most advanced video coding algorithm that the ITU has standardized.

Quantization refers to the step in video coding in which the DCT coefficients are subject to integer rounding. Subsequent to quantization, these quantized DCT coefficients are run-length coded, which involves coding runs of zeros present. This is achieved by using a variable length code to code events of the form (LEVEL,RUN,LAST), where LEVEL is the value of the quantized coefficient immediately following the run of zeros, RUN is the length of the run of zeros, and LAST is a flag indicating that the current coefficient is the last nonzero coefficient in the block. Provided that the H.263+ syntax is obeyed, the decisions on how to map input DCT coefficients c_i into quantized coefficients q_i at the encoder are left to the implementer, and it is this flexibility that opens the door to rate distortion optimizations.

The quantization process can be represented using a trellis as shown in Figure 1. The stage i in the trellis corresponds to the ith DCT coefficient. For each stage, there are a total of J LEVEL states and K RUN states. The jth LEVEL state of the ith stage, denoted by $L_j(i)$, corresponds to the action of quantizing the ith coefficient to LEVEL j; the kth RUN state in stage i, $R_k(i)$, corresponds to the case where the ith coefficient is quantized to zero **and** that the previous $k-1$ coefficients were also quantized to zero. Thus, the cost for including the dependencies between coefficients is reflected in the need to have enough RUN states to allow representation of the maximum possible run length.

As with the Viterbi algorithm, the quantization algorithm relies on the concept of surviving paths. However, in contrast with the Viterbi algorithm in which one metric (distortion) is considered, the trellis in Figure 1 is being used to consider both rate and distortion. The algorithm fills in the trellis by retaining, for each state, the incoming path with the lowest accumulated $R + \lambda D$ metric. When the trellis has been processed through all 64 stages, it becomes possible to compute the rates associated with all paths entering RUN states, as well as the LEVEL states, and the complete metric is known for all incoming paths. In addition, the metric is computed for the case in which all coefficients are quantized to zero, in which case the associated rate is zero. If this is the minimum metric, this block will not be coded by signalling with the corresponding H.263+ syntax. Otherwise, the state in the final stage with minimum incoming path metric is chosen, and its surviving path is followed back to the beginning of the trellis to determine the quantized coefficients.

3 Experimental Results

Experiments were performed on blocks of 11 frames, the first compressed using purely INTRA techniques and the remaining frames compressed using INTER coding. The bit rates ranged from 20.76kbps for the low-motion "claire" sequence to 49.59kbps for the high-motion "foreman" sequence. Table 1 gives results with PSNR of the luminance component measured in the image domain. The average bitrate savings is approximately 3.5% for equivalent image noise for the carphone and foreman sequences. The improvement for the Claire sequence is limited by the low motion, and there is no improvement in bit rate.

4 Conclusions

We have presented an algorithm which performs the R-D optimal quantization of DCT coefficients in the H.263+ video coding system including run length and Huffman coding. With a fixed PSNR, the R-D approach reduces the bitrate by approximately 3.5%, or equivalently with a fixed bitrate, PSNR performance is improved by an average of 0.17dB over the carphone and foreman sequences.

While the goal of this work was to establish the benefits of R-D optimized quantization with complexity as a secondary consideration, complexity is certainly an issue in practical implementations. Running on a Sun Ultra Sparc 2 workstation with 64 megabytes of RAM, the R-D optimized approach runs about 30 times slower than the publicly available H.263+ coder implementation from the UBC [6]. However, a fast algorithm exists with substantially lower processing requirements. Since 95% of coefficients are quantized to 0 with the standard quantization equations, eliminating their corresponding stages from the trellis would eliminate 95% of the trellis processing time, leading to an execution time only 1.5 times that of the UBC coder.

5 Acknowledgement

The authors would like to thank Gary Sullivan for his insightful comments regarding this work.

References

[1] ITU-T (formerly CCITT), "Video Coding for Low Bit Rate Communication," ITU-T Recommendation H.263; version 1, Nov. 1995; version 2, Jan. 1998.

[2] S. Wu, A. Gersho, "Rate-constraint picture-adaptive quantization for JPEG baseline coders," *Proc. Inter. conf. ASSP*, pp. 389-392, Apr. 1993.

[3] M. Crouse, K. Ramachandran, "JPEG optimization using an entropy-constraint quantization framework, " *Proc. IEEE Data Compression Conference*, pp. 342-351, Snowbird UT, 1995.

[4] K. Rijkse, "H.263: Video Coding for Low-Bit-Rate Communication", *IEEE Communications Magazine*, pp. 42-45, Dec. 1996.

[5] T. Sikora, "MPEG Digital Video-Coding Standards", *IEEE Signal Processing Magazine, vol. 14, no. 5*, pp. 82-100, Sep. 1997.

[6] M. Gallant, G. Cote, B. Erol, F. Kossentini, "UBC's H.263+ Public Domain Software, Release 1", ITU-T Document Q15-B-33, Sunriver, Oregon, Sep. 1997.

Sequence	Quantizer	PSNR(dB)	Bitrate(kbps)
Foreman	Reference	30.68	49.59
	Trellis	30.84	49.37
	Trellis	30.72	47.67
Claire	Reference	34.82	20.76
	Trellis	34.84	20.75
Carphone	Reference	31.73	39.24
	Trellis	31.91	39.15
	Trellis	31.77	38.04

Table 1: Comparison of reference and trellis-based codecs at identical bitrate and PSNR

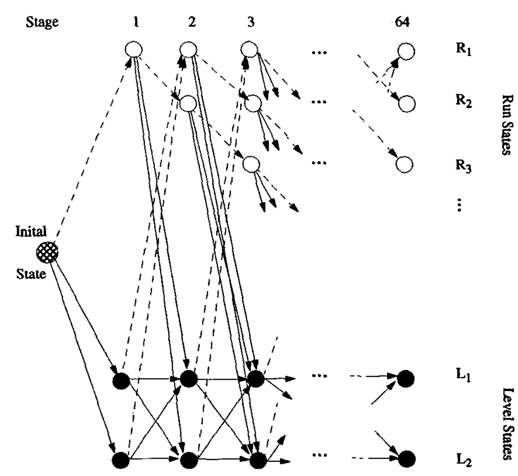

Figure 1: Block DCT quantization trellis