Syntax-constrained encoder optimization using adaptive quantization thresholding for JPEG/MPEG coders

Kannan Ramchandran
Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign

Martin Vetterli
Department of Electrical Engineering and Computer Science
University of California at Berkeley

January 17, 1994

Abstract

We show a rate–distortion optimal quantization technique to thresh-
old the DCT coefficients in the industry image and video coding stan-
dards JPEG and MPEG respectively. Our scheme achieves a decent
thresholding gain in terms of both objective SNR (about 1 dB) as
well as perceived quality and uses a fast dynamic programming recur-
sive structure which exploits certain monotonicity characteristics of
the JPEG and MPEG codebooks to drastically reduce the complex-
ity. The primary advantage of our encoding algorithm is that it is
completely compatible with the baseline JPEG and MPEG decoders.

1 Introduction

Recent advances in digital image and video compression techniques have
culminated in the formulation of the popular CCITT standards for still–
image and video compression formats, JPEG [1] and MPEG [2] respectively.
Although the *syntax* of these coding formats is constrained by the proposed
standards, there remains considerable freedom to optimize the coder while
being syntax-constrained. This leads to the paradigm of decoder-compatible
encoder optimization, which has obvious applications to the optimal use of
"existing" decoders. In this paper, as an extenion of our work in [6], we
address decoder-compatible encoder optimization applied to the JPEG and

MPEG coders, with the obvious advantage being that the user does not have to change his existing decoder, an important practical consideration in light of the ever-increasing volume of existing JPEG and MPEG decoders in the market today.

It must be pointed out that though this paper is directed at the optimization of the JPEG and MPEG coding formats, its scope of application (i.e. the use of quantization thresholding) is broad, and encompasses general coding systems. In particular, wavelet based coding methods have received considerable attention in the literature lately due to their good performance [7]. The use of quantization thresholding methods to the wavelet coder has recently been applied to a wavelet-based coder with considerable success [9, 10].

JPEG and MPEG are based on the Discrete Cosine Transform (DCT). Non–overlapping 8x8 blocks of the image or video frame are 2–D DCT transformed, quantized using a perceptually weighted quantization matrix, scanned in a zigzag order, and subjected to a 2–D variable length code (VLC) using Huffman coding tables. Refer to [1, 2] for details. The widely–used "default" Huffman coding tables of JPEG and MPEG are derived from statistics of typical training images and video sequences respectively, and are not customized for arbitrary inputs. While the standards allow for custom Huffman codebooks for each input source, this represents, in most cases, an expensive overhead (as the Huffman table has to be sent). This therefore provides the motivation to investigate ways of introducing adaptivity of the coding tables *without actually changing the default codebook,* i.e. without incurring the overhead of sending the customized codebook information.

A key to good compression when using these standards *while still being compatible with the standard decoder* lies in sending "smart" subsets of the transform coefficients used to represent the image or video sequence. This could be viewed as an application of entropy-constrained quantization [4] to the existing JPEG/MPEG codebooks. Thus, from the existing set of all possible quantization levels that can be applied to the transform coefficients, we are interested in picking the choice which minimizes the distortion subject to the rate constraint, as dictated by the use of the JPEG/MPEG codebooks.

Thus, our goal consists in determining which DCT coefficients are "significant" enough to be coded and which ones are better off being thresholded or dropped. Dropping or thresholding of the less significant coefficients may be desirable in the rate–distortion sense, as it may lead, at a marginal sacrifice of coded quality, to a significant reduction in expended bit rate, due to fewer coefficients needing to be transmitted. This is especially so when

deciding the last non–zero coefficient, which is typically followed by an inexpensive "end–of–block" code [1, 2]. Thus, it may especially pay to get rid of compression–hindering sparsely interspersed insignificant coefficients that represent the last non–zero values before the end–of–block. In the case of JPEG, where the standard does not permit variable scaling of the different image blocks, intelligent adaptive thresholding of the coefficients would tantamount to changing the quantization scales at the block level, without violating the syntax rules.

We formulate an optimal strategy to drop or threshold the quantized DCT coefficients using a fast recursive technique employing dynamic programming (DP). While in [6], we restricted ourselves to "hard" thresholding (i.e. making a binary decision on keeping or dropping a coefficient), in this work, we consider "soft" thresholding methods also (i.e. we allow for coarser but non-zero quantization while being syntax-compatible). The optimality is in a rate–distortion (R–D) sense, for the fixed scale quantization performed prior to thresholding. Thus, starting from the "highest quality point" after quantization at a fixed scale (for JPEG) or QP–level (for MPEG [2]), one can sweep the entire thresholding R–D curve over a continuous range of target bit rates (or equivalently target coding qualities) by thresholding insigificant coefficients in the image or video frame. Thus, our algorithm could find all points that reside on the convex hull [1] [3] of the R–D curve. The appeal of the strategy lies in its combination of R–D optimality, speed of operation, and its complete compatibility with standard JPEG and MPEG decoders, which remain blissfully oblivious to the thresholding gymnastics performed by the encoder. A point to note is that the algorithm described in this work exploits the monotonic nature of bit rate versus the zero run–length count preceding a non–zero coefficient inherent in the Huffman tables of JPEG and MPEG, to specify a fast "pruning" operation in the DP recursion. Thus, the computational complexity of the algorithm (at least for the hard thresholding case) becomes low enough for it to be implementable.

This paper is organized as follows. Section 2 states the problem formally, while Section 3 gives the basic idea of the optimal algorithm. Coding applications using JPEG and MPEG are presented in Section 4.

---

[1]i.e. the convex boundary of R-D points (see Fig. 3)

2 Problem statement

We wish to find that optimal set of post–quantization DCT coefficients to be retained for every 8x8 block of an image or video frame such that the mean–squared–error (MSE) distortion (any additive distortion metric is feasible in general, e.g. weighted MSE) between the original image and the thresholded version is minimized *subject to a maximum target coding bit rate constraint* or equivalently, the coding bit rate is minimized subject to a maximum allowable distortion constraint.

See Figure 1. Thus, if X is the input signal, $\hat{X}$ the quantized output corresponding to a fixed scale or "anchor" level representing the maximum quality operating point, and $\tilde{X}$ the thresholded version of $\hat{X}$, we seek to minimize the MSE distortion between X and $\tilde{X}$ given the quantized image $\hat{X}$, subject to a total coding bit budget of R_{budget} for $\tilde{X}$. That is, our goal is to find:

$$\min[\mathcal{D}(X,\tilde{X})|\hat{X}] \text{ subject to } \mathcal{R}(\tilde{X}) \le R_{budget}. \tag{1}$$

3 Basic idea of solution

R–D optimality: the constant slope condition

The "hard" constrained thresholding problem of Eq.(1) can be solved by being converted to an "easy" equivalent unconstrained problem by "merging" rate and distortion through the Lagrange multiplier λ [8]. The unconstrained thresholding problem becomes the determination (for a fixed λ) of that set of coefficients, which results in the minimum total Lagrangian cost defined as $J(\lambda) = D(X,\tilde{X}) + \lambda R(\tilde{X})$.

The optimal coefficient search for the image *can be done independently* for every 8x8 image block for the fixed quality "slope" λ, which trades off distortion for rate. This is because it can be shown [8, 5] that, at R–D optimality, all blocks *must operate at a constant slope point λ on their R–D curves*. The desired optimal constant slope value λ^* is not known *a priori* and depends on the particular target budget or quality constraint. Fortunately, however, the search for λ^* is a fast convex one, and it can be obtained using a fast convex recursion in λ using the bisection algorithm [5].

Fast dynamic programming algorithm

Since the optimal solution can be obtained by finding the minimum-Lagrangian-cost operating point for each block independently, it suffices to consider a single block for analysis. The problem is solved using a dynamic programming approach.

The zigzag scan that is part of the standards [1, 2] is used to order the 2-D coefficients. As an initialization, one has to gather the $\Delta J_{j,k}$'s associated with the incremental Lagrangian cost of going from coefficient j to coefficient k (i.e. dropping all the coefficients between them) for all non–zero valued (j, k) coefficient pairs with $j < k$. The optimal operating point, for a fixed value of λ, can be done in a recursive fashion by finding (i) the minimum Lagrangian cost, $J^*(k)$ and (ii) the optimal predecessor coefficient, *predecessor(k)*, associated with choosing coefficient k as *the last non–zero coefficient* for all $k = 1, 2, \ldots, 63$. Then, starting from that coefficient k^* which is the cheapest to retain as the last non–zero coefficient, i.e. minimum $J^*(.)$, the optimal set can be "backtracked" from the optimal predecessor chain calculated for all predecessors of k^*. See Figure 2.

In finding the optimal predecessor at a particular iteration k as described above, one generally has to consider as candidates ALL coefficients $i < k$. However, for the particular case of monotonicity of $R_{i,k}$ in the zero runlength count $(k - i - 1)$, which is true for the coding tables of JPEG and MPEG, a fast pruning algorithm can be used to speed up the search. This leads to a substantial decrease in computational complexity, and leads to a fast optimal algorithm. The above optimal DP algorithm is performed independently on all blocks. The composite R–D point for the picked λ is obtained simply as the sum of the optimally obtained R–D points for each block for that λ. Finally, the optimal slope λ^* which solves the desired budget or quality constraint is found with a fast convex search using the bisection algorithm [5].

Soft thresholding

While the above discussion describes the optimal algorithm for doing "hard" thresholding –i.e. in reaching a binary decision on whether or not to drop a coefficient, it can be modified to include soft decisions as well. Thus, it may pay, in the rate-distortion sense, to quantize some coefficients coarser but not zero them out altogether. Thus, the hard thresholding algorithm would become a special case of the soft thresholding paradigm, though the latter has

significantly more computational complexity. Experiments of the soft thresh-olding technique are ongoing, though preliminary results indicate that the benefit of using soft thresholding for JPEG/MPEG applications is probably unprofitable in the complexity-performance tradeoff sense, as we are dealing with codebooks which are optimized for zero runlengths. However, the gain of applying soft thresholding techniques to other coding environments which are not zero run-length coders (e.g. the wavelet-based entropy coder of [10] may be significant.

4   Application

The optimal algorithm outlined in Section 3 can be used to quantify the benefits of adaptive thresholding applied to the JPEG and MPEG coding environments. R–D curves are obtained by sweeping the Lagrange multiplier λ through all positive values for typical quantization scales of interest. Figure 3 shows the R–D curves for a typical image ("House") using JPEG for pre-thresholding quantization scales of 1.0 and 0.7. Note the significance of Figure 3. Point X on curve (a) is the unthresholded "reference" obtained for a scale of 1.0. Let us fix the bit rate for the problem at the reference X's bit rate of 0.615 bits per pixel (bpp). Now, to see the advantage of optimal thresholding, observe curve (b) corresponding to a finer quantization scale of s=0.7 instead of 1.0. For this finer scale, the non–thresholded bit rate, corresponding to point Z, is obviously greater than that of X. However, if we now start thresholding optimally until the bit budget constraint imposed by curve (a) is satisfied, *we will get an adaptive thresholding gain in terms of increased SNR for the same bit rate.* Thus, point Y enjoys a 0.7 dB gain at the same bit rate over X. Alternatively, if we fix the PSNR according to that of X (35.17 dB), point W enjoys a compression advantage at the same PSNR of roughly 15 %. Coding results show that for typical images and video sequences, a gain of 0.5-0.8 dB is obtainable while being completely compatible with the decoder!

References

[1] JPEG technical specification: Revision (DRAFT), joint photographic experts group, ISO/IEC JTC1/SC2/WG8, CCITT SGVIII, August

$$X \longrightarrow \boxed{\text{QUANTIZER}} \overset{\hat{X}}{\longrightarrow} \boxed{\text{THRESHOLDER}} \longrightarrow \check{X}$$
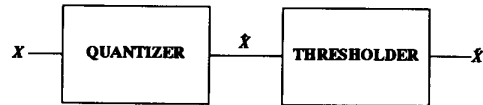
PROBLEM: $\min\limits_{\{\text{all admissible } \check{X}\}} D(X, \hat{X}|\check{X})$ subject to $R(\check{X}) . R_{budget}$

Figure 1: Block diagram of problem statement: X is the input image, $\hat{X}$ is the quantized pre–thresholded version (after JPEG or MPEG), and $\check{X}$ the output of the thresholder. We want to minimize the distortion, for a fixed quantizer, between the original and the thresholded images subject to a bit budget constraint.

1990.

[2] MPEG video simulation model three, ISO, coded representation of picture and audio information, 1990.

[3] P. A. Chou, T. Lookabaugh, and R. M. Gray. Optimal pruning with applications to tree-structured source coding and modeling. *IEEE Trans. on Inform. Theory*, IT-35:299–315, March 1986.

[4] P. A. Chou, T. Lookabaugh, and R. M. Gray. Entropy-constrained vector quantization. *IEEE Trans. on Acoust., Speech, Signal Proc.*, 37(1):31–42, January 1989.

[5] K. Ramchandran and M. Vetterli. Best wavelet packet bases in a rate-distortion sense. *IEEE Trans. on Image Proc.*, 2(2):160–175, April 1993.

[6] K. Ramchandran and M. Vetterli. Rate–distortion optimal fast thresholding with complete JPEG/MPEG decoder compatibility. In *PCS'93, Lausanne, Switzerland*, 1993.

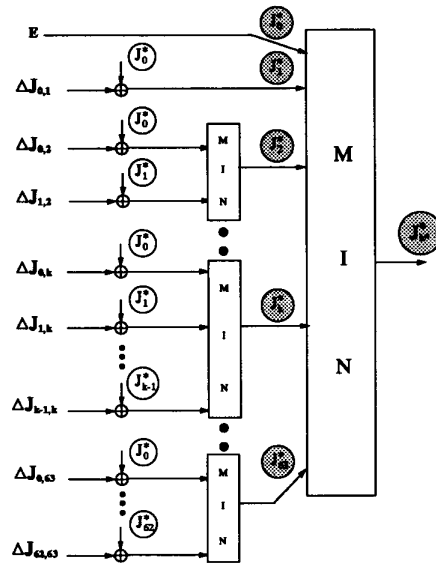[7] J. M. Shapiro. An embedded wavelet hierarchical image coder. *Proc. of ICASSP*, pages 657–660, March 1992.

Figure 2: Block diagram of the dynamic programming recursion of the optimal algorithm for each 8x8 image block. E is the A.C. energy, $\Delta J_{i,j}$ is the incremental Lagrangian cost of going from coefficient i to coefficient j while dropping all coefficients in between, and J_k^* is the minimum Lagrangian cost associated with ending at coefficient k. Note that $k^* = \min_{0 \leq k \leq 63}^{-1}$ is the optimal "last" coefficient.
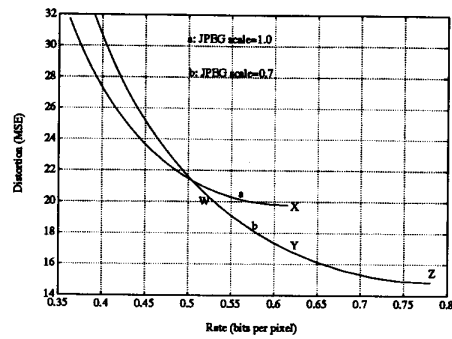
Figure 3: Optimal thresholding R-D curves for "House" image using JPEG. Curve (a) corresponds to a JPEG quantization scale of 1.0, while curve (b) to a finer scale of 0.7. Note that if we fix the reference at point X on curve (a) corresponding to a scale of 1.0, we can achieve point Y by "backing off" to a finer scale of 0.7 (point Z on curve b) and then thresholding optimally to point Y at the same bit rate as X. Note that the thresholding gain for this example (reference X) is approximately 0.7 dB at a bit rate of 0.615 bpp (point Y), or alternatively about 15% reduction in bit rate (point W of curve b) at the same PSNR of 35.17 dB.

[8] Y. Shoham and A. Gersho. Efficient bit allocation for an arbitrary set of quantizers. *IEEE Trans. on Acoust., Speech, Signal Proc.*, 36(9):1445–1453, September 1988.

[9] Z. Xiong, N. P. Galatsanos, and M. T. Orchard. Marginal analysis prioritization for image compression based on a hierarchical wavelet decomposition. In *Proc. of ICASSP'93*, pages 546–549, April 1993.

[10] Z. Xiong, K. Ramchandran, and M. T. Orchard. Joint optimization of scalar and tree-structured quantization of wavelet image decomposition. In *Asilomar conference on signals, systems and computers*, November 1993.