OPTIMISATION OF TWO-LAYER SNR SCALABILITY FOR MPEG-2 VIDEO

D. Wilson and M. Ghanbari

Department of Electronic Systems Engineering,
University of Essex, Colchester CO4 3SQ, UK.
E-mail: dwilso@essex.ac.uk, ghan@essex.ac.uk

ABSTRACT

SNR scalability, used in two-layer video coding, guarantees good base quality pictures at the expense of increased overall bit-rate. By understanding the inherent inefficiencies of enhancement layer coding we have developed an optimisation method called *optimal coefficient adjustment* in order to reduce overall bit-rates to levels consistent with single-layer operation.

1. INTRODUCTION

SNR scalability (SNRS) has been proposed by MPEG-2 in order to address some of the problems presented by applications such as ATM networks and terrestrial broadcasting, where a minimum base quality is to be maintained in the event of information loss. The advantage of SNRS is its ability to provide superior base quality to that of the alternative method of data partitioning (DP) for a given equivalent base bit-rate. Coarse quantisation performed by the base effectively filters out insignificant DCT coefficients with those remaining providing the essential structure of the picture and requiring fewer bits. In contrast, the insensitivity of DP to coefficient magnitude often prevents important higher frequency content from being included in the base layer.

The main drawback of SNRS is a significant increase in bit allocation which may generate as much as 10–15% more bits than its single-layer equivalent operating at the same quantisation index [1], while DP incurs only a small increase in bit-rate attributable to the duplication of slice header information, typically resulting in only a 1–2% increase in bit-rate.

In the past few years several optimisation techniques have been applied to video coding in order to improve coding efficiency in a rate-distortion context [2, 3]. More recently the optimal *thresholding* technique presented in [3] has been adapted and applied specifically to SNRS by De Lameillieure [4] demonstrating significant improvement in enhancement layer picture quality over non-optimised coding operating at the same constant bit rate. In this paper we present a novel method of *optimal coefficient adjustment* that extends the concept of coefficient thresholding to include the possibility of 'trimming' rather than 'dropping' quantised DCT coefficients. Such an approach, it shall be shown, allows a much finer control over rate-distortion resulting in further gains in terms of bit-allocation for a given picture quality.

The paper is structured as follows: in section 2. we explain the SNRS coder and identify inherent coding deficiencies of the enhancement layer; based on this, in section 3. we introduce our proposed *optimal coefficient adjustment* method aimed at reducing enhancement layer overheads; section 4. describes experimental simulations carried out in order to assess performance. Finally section 5. concludes the paper.

2. SNR SCALABILITY

MPEG-2 defines SNR scalability as a mechanism to refine (or *enhance*) the DCT coefficients encoded by another layer [5]. In the decoder the reconstructed 8×8 DCT coefficients are formed by combining the inverse quantised coefficients of the base and enhancement layers.

Since the MPEG-2 standard permits a degree of flexibility within the encoder, motion compensation may either be performed based on base or enhancement reconstructed picture qualities; if based on base quality an accumulation of discrepancies or *drift* between encoder and decoder pictures will occur when the decoder includes the enhancement layer; if based on enhancement quality drift will occur only in the absence of the enhancement layer; Either is permitted, but as enhancement should provide the best possible quality it is appropriate to opt for the latter, especially as this reduces the overall bit-rate [6].

In this case, we can define our *standard* SNR scalable encoder as illustrated in Fig. 1. The base layer is formed first by quantising the 8×8 DCT coefficient block using a coarse quantisation step size q_b which is then variable length coded to form the base bit stream. The reconstructed (inverse quantised) base block is then subtracted from the original DCT coefficients in order to derive the *residual quantisation error* (RQE) introduced by the base. RQE is then finely quantised using the enhancement quantisation index q_e and variable length coded, in the same way as the base layer, to form the enhancement bit stream.

The quantisation level applied to the differentially coded DC coefficient of intra-coded blocks is fixed independently of base and enhancement quantisation indices and hence is coded once in the base layer and should not be recoded in the enhancement layer. The DC coefficient does not therefore contribute to SNRS inefficiency. For this we must look to the remaining 63 AC coefficients plus the DC coefficient of inter-coded blocks. These are quantised according to the

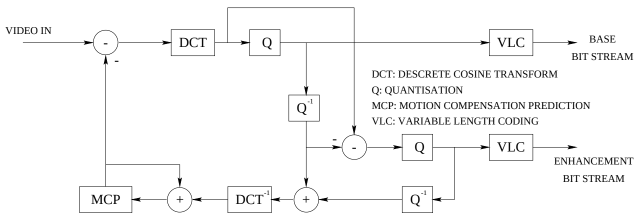

Figure 1. ***Block diagram of SNR scalable coder***

quantisation indices, q_b and q_e, which are fixed for each macro-block and which result in a quantisation integer step sizes, s_b and s_e, where s is defined as [5]:

$$s = \frac{w \times q}{8} \tag{1}$$

with weighting factor w predefined according to coefficient position in the 8×8 transform matrix.

In order to achieve better coding compression, quantised coefficients are also thresholded in order to remove less important DCT coefficients. It is the resulting *dead-zone* in the enhancement layer that is responsible for the slight degradation in enhanced picture quality that is characteristic of SNRS coders which may be as much as 0.1–0.2 dB PSNR less than the equivalent single-layer.

In addition to the same header duplications incurred by DP, SNRS is also subject to bit allocation inefficiencies arising from the variable length coding (VLC) of the enhancement layer; caused by the following factors:

i.) run-lengths, used to index VLC code tables, are in general longer for SNRS coders, due to the absence of certain non-zero coefficients that are encoded only in the base and others that are encoded only in the enhancement layer, effectively leaving 'gaps' in both layers;

ii.) many coefficients which appear both in the base and enhancement layer may, due to the non-linear nature of the VLC tables, have a combined bit allocation that exceeds that required to code the same coefficient in an equivalent single-layer coder;

iii.) the reduced set of enhancement layer quantised integer values is no longer matched to the VLC tables which are based on the distribution of a full range of quantised DCT values. Individually tailoring an enhancement layer VLC table corresponding to each base/enhancement quantisation index pair will in most cases be impractical.

The extent to which enhancement picture quality is degraded and bit allocation efficiency is compromised is dependent on the value of s_b relative to s_e, and therefore according to (1), q_b relative to q_e; the larger q_b is relative to q_e the smaller the proportion of RQE falling within the enhancement dead-zone and also the greater the number of coefficients to be coded only in the enhancement layer. In other words the coarser the base layer the closer SNRS performance approaches that of its single-layer equivalent. Nonetheless, a minimum base quality may dictate enhancement layer efficiency which will for most applications be constrained by, say, the degree of 'gracefulness' to be provided by the base layer. Thus, bit-rate overheads of up to 10–15% [1] of equivalent single-layer bit-rates may in some cases be unavoidable — unless steps are taken to reduce them.

3. MODIFYING RATE-DISTORTION USING DCT COEFFICIENT ADJUSTMENT

Although, the optimisation of the quantisation indices, q_b and q_e, may be performed using standard Lagrangian techniques [7], optimisation of the quantisation indices alone, whilst of benefit, is not sufficient to address the inherent enhancement layer coding inefficiencies of SNRS which tend to be present regardless of index. Thus, what we require is a method to optimise bit allocation between base and enhancement layers of each quantised DCT block.

The method of selectively thresholding certain coefficients in order to optimise rate-distortion in single-layer DCT block coding has already been presented by Ramchandran and Vetterli [3]. Of particular interest was the method that optimally thresholded coefficients of a finely quantised block, operating with index q', to achieve the target distortion corresponding to a coarser index, q, in order to achieve a lower bit-rate. As described by De Lameillieure [4], the same process may be applied SNRS coders in one of two ways: as two separate 1 D processes performed first on the base and then on the enhancement layers; or as a combined 2 D process. The results presented in [4], suggest the more complicated 2 D approach delivers only a small improvement over separate base and enhancement 1 D approach. Furthermore, our own experiments have shown that it is the optimisation of the enhancement layer, the source of SNRS bit allocation inefficiencies, that delivers the largest gains. Complexity can therefore be reduced, at a negligible cost to performance by only applying optimisation to the enhancement layer.

Unlike the base layer, the range of quantised coefficient magnitudes in the enhancement layer is limited to

$$0 \leq |z| \leq |Q_e(RQE_{max})|, \tag{2}$$

for each quantised coefficient z, where $|Q_e(RQE_{max})|$ is defined by base dead-zone and step size, s_b. Because of this, and also the resulting longer run lengths, the enhancement layer tends to be much more sensitive to thresholding than its single-layer counterpart.

As thresholding can be considered as being a small subset of a more general approach to coefficient manipulation, the search for the optimal DCT block combination can be improved, in a rate-distortion context, by allowing enhancement layer DCT coefficients to be *adjusted* over the full range of z. Thus, consider the quantised coefficient, z, having been quantised at the finer quantisation index, q'. Instead of either 'keeping' or 'dropping' z, we may 'adjust' z by an integer amount u resulting in a modified coefficient, $|z'|$, where:

$$|z'| = |z| - u. \tag{3}$$

The magnitude of quantised coefficients can therefore be 'trimmed', trading distortion against bit allocation without suppressing completely the presence of any given DCT coefficient.

In addition, the adjustment value u may also be negative, such that $z' > z$. There is, in general, very little reason for doing this since not only do we increase distortion but also bit allocation without any obvious benefit. However, the case where $z = 0$ presents one important exception where the introduction of $z' = 1$ (i.e. $u = -1$) can be beneficial in breaking up critically long zero run lengths to the next non-zero coefficient in order to avoid, say, large escape codes. Indeed, the presence of the dead-zone means that certain $z = 0$ coefficients can be replaced by $z' = 1$ and at the same time reduce distortion. The search range of potential coefficient adjustments can thus be bounded by the limits

$$-1 \leq u \leq |z|. \tag{4}$$

Given a group of N blocks, for which the quantisation indices have been predetermined, we are now faced with the task of finding the optimal coefficient adjustment for each of the $63 \times N$ AC coefficients. Our aim is to minimise bit allocation subject to a minimum total distortion target $\tilde{D}$, or conversely, to minimise distortion subject to a total bit allocation target $\tilde{R}$. Here, we shall concentrate on achieving the former, $\tilde{D}$, for two reasons; a.) for VBR applications it is easier to specify *a priori* a minimum desired quality, and b.) our aim is to reduce SNRS bit-rate whilst maintaining equivalent distortion so that we might compare the results with equivalent coders operating at the same quantisation indices.

The total distortion, D, for N blocks is defined as the sum of the square quantisation errors, contributed by each of the $63 \times N$ AC coefficients, for any given adjusted coefficient combination. If $d'_{i,j}$ represents the distortion, $x_{i,j}$ the original DCT coefficient, and $\hat{x}'_{i,j}$ the adjusted and reconstructed DCT coefficients for the j^{th} coefficient in the i^{th} block, the total distortion, D, is given by:

$$D = \sum_{i=1}^{N} \sum_{j=1}^{63} d'_{i,j} = \sum_{i=1}^{N} \sum_{j=1}^{63} (x_{i,j} - \hat{x}'_{i,j})^2 \tag{5}$$

Given a target distortion, $\tilde{D}$, the problem can be specified succinctly as

$$\min\{R(z')\} \qquad \text{subject to} \qquad D(x, z') \leq \tilde{D}, \tag{6}$$

a constrained problem that may be readily solved by transforming it into an unconstrained problem by merging rate and distortion using the Lagrangian multiplier λ to produce a Lagrangian cost

$$J_{min}(\lambda) = \min\{D(x, z') + \lambda R(z')\}, \tag{7}$$

where λ is found iteratively exploiting the convex nature of the rate-distortion curve to converge quickly (typically 8–10 iterations) to a solution.

The number of permutations P_{adj} corresponding to each combination of adjusted coefficients for any given block is defined as

$$P_{adj} = \prod_{k=1}^{63} (z_k + 2) \tag{8}$$

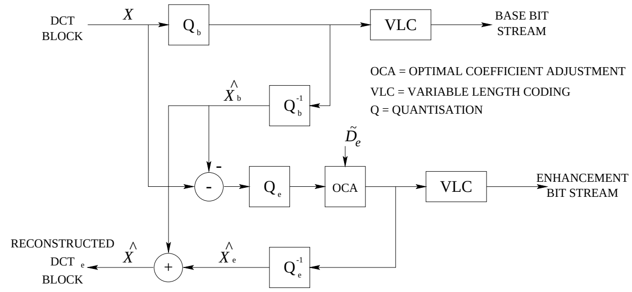

Figure 2. ***Block diagram of optimised SNR quantisation stage***

where according to (4) each pre-adjusted coefficient, z_k, can be adjusted in $z_k + 2$ ways.

Each combination of a block represents one point in the rate-distortion plane with all the combinations together forming a 'cloud' of R-D points whose lowest outermost minimum R-D points describe a convex curve of optimum solutions. All points that exist above this curve are sub-optimal. A brute force calculation of every Lagrangian cost J, corresponding to each P_{adj} combination, can be avoided by exploiting DCT block coding characteristics, and in particular the *monotonicity* property of of MPEG-2 VLC tables, in order to *prune* sub-optimal combinations at an early stage in the calculation. So doing, the number of 'surviving' combinations after calculating J for the k^{th} coefficient is reduced from (8) to simply

$$2 \leq P_{adj} \leq k + 1. \tag{9}$$

There are thus between 2 and 64 surviving combinations possible for each block of the 63 AC coefficients, from which the final combination with the minimum overall cost can be selected.

4. EXPERIMENTS

A pre-recorded video sequence was used in order to compare our optimal coefficient adjustment method with both standard SNRS and single-layer equivalents. Fixed quantisation indices are used in order to facilitate comparison between coders resulting in bit streams that are inherently VBR. Nevertheless, it should be noted that the same process is equally applicable to CBR operation.

The enhancement layer is optimised according to the modified quantisation stage shown in Fig. 2. Optimisation is performed over each MPEG-2 slice (i.e. $N = 22 \times 6$ blocks), although macro-block, frame, GOP or other intervals are also valid depending on delay constraints, and/or buffering and processing capacities.

Bit-rate reductions are measured relative to the standard SNRS process operating with fixed indices, q_b and q_e, such that enhancement layer distortions of the standard SNRS process are used as the target distortion, $\tilde{D}_e$, for the optimised SNRS coder operating with indices q_b and $q'_e = q_e - \alpha$ ($\alpha > 0$); where α is chosen to maximise coding gain.

Fig. 3 shows rate-distortion curves for the enhancement layer of a typical MPEG-2 coded slice, to which several enhancement quantisation indices are applied whilst maintaining a constant base index of $q_b = 8$. The 'stand-

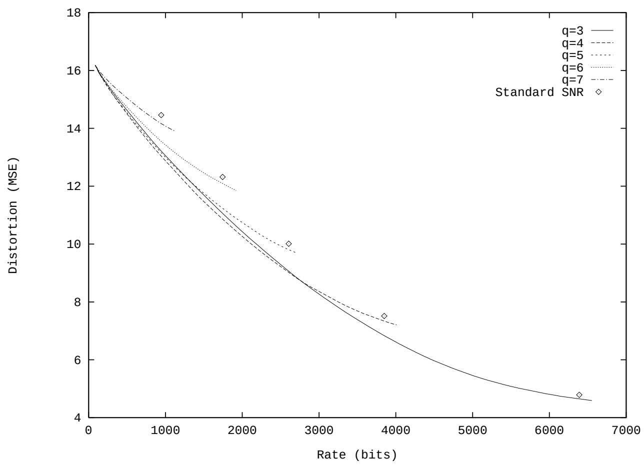

Figure 3. ***Enhancement layer R-D curves*** ($q_b = 8$)

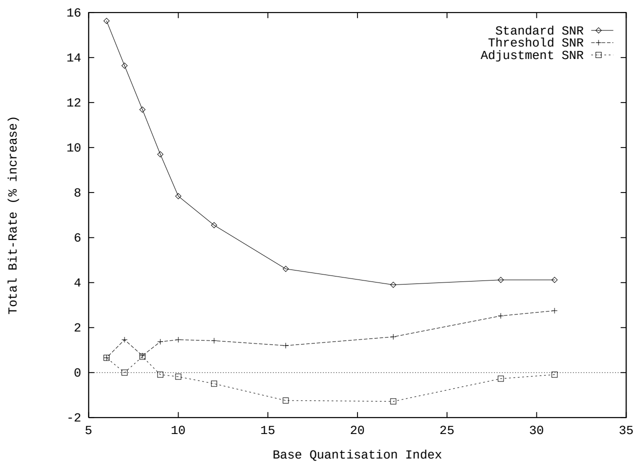

Figure 4. ***Total SNRS bit-rate (relative to single-layer) vs*** q_b.

ard SNR' points shown accompanying correspond to the non-optimised SNRS coder operating at the same range of indices (i.e. q_e=3–7). It can be seen that each enhancement quantisation index is clearly able to outperform its coarser neighbours with significant sections of coarser curves lying to the right of finer ones.

To provide a reference point, the pre-recorded test sequence was single-layer coded with quantisation index arbitrarily set to $q = 5$, producing a VBR bit stream with an average bit-rate of approximately 2MB/s. The enhancement layer index of the standard SNRS is thus also fixed at $q_e = 5$ to be used as the target distortion, $\tilde{D}_e$, for the optimised SNRS coder whose enhancement layer index is set to $q'_e = 5 - \alpha$. The combined base and enhancement bit-rate is then measured against a range of base quantisation indices between the permissible limits $q_e < q_b \leq 31$.

The results are presented in Fig. 4 which shows a graph of standard and optimised SNRS total bit-rates, versus base quantisation index. Total rates are presented as a percentage increase over the single-layer equivalent in order to give some idea of 'cost' incurred to replace one layer with two.

The graph shows how standard SNR scalability performance degrades as q_b approaches q_e, such that at its worst ($q_b = 6, q_e = 5$), the standard SNRS produces a bit-rate over 15% higher than its single-layer equivalent, as well as 0.2 dB drop in enhancement layer PSNR. Yet, even when q_b is at its coarsest, standard SNRS coding still results in a total bit-rate requiring 4% more bits than the single-layer.

Both optimal thresholding and adjustment methods can be seen to substantially reduce SNRS overheads. While not shown here, enhancement layer quality is also slightly improved helping to restore quality loss due to requantisation. For finer q_b, where most coefficients can only be zero or one, the difference between thresholding and adjustment is nominal; the only benefit of adjustment being the ability to introduce non-zero coefficients. As q_b becomes coarser, the integer range of quantised enhancement coefficients increases along with the scope for adjustment enabling optimal adjustment to outperform thresholding by a more significant margin. Furthermore, it can be seen that only optimal coefficient adjustment is able to reduce SNRS bandwidths to below that required by the single-layer equivalent.

5. CONCLUSIONS

SNRS performance has been shown to benefit significantly from optimisation. In particular, there is much scope for improving enhancement layer coding which has been identified as being the source of SNRS bit allocation inefficiencies. We have described optimal coefficient adjustment and demonstrated its capability to outperform existing thresholding techniques.

REFERENCES

[1] D. Wilson and M. Ghanbari, "Transmission of SNR scalable two-layer MPEG-2 coded video through ATM networks," in *7th Int. Workshop on Packet Video*, (Brisbane, Australia), pp. 37–42, March 1996.

[2] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," *IEEE Transactions on Accoustics, Speech, and Signal Processing*, vol. 36, pp. 1445–1453, September 1988.

[3] K. Ramchandran and M. Vetterli, "Rate-distortion optimal fast thresholding with complete JPEG/MPEG decoder compatibility," *IEEE Transactions on Image Processing*, vol. 3, pp. 700–704, September 1994.

[4] J. De Lameillieure, "Rate-distortion optimal thresholding in SNR scalability based on 2D dynamic programming," in *SPIE Proceedings: Advanced Imaging and Network Technologies*, vol. 2952, pp. 689–698, October 1996.

[5] ISO/IEC, "Generic coding of moving pictures and associated audio, recommendation H.262," March 1994.

[6] M. Ghanbari and V. Seferidis, "Efficient H.261-based two-layer video codecs for ATM networks," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 5, pp. 171–175, April 1995.

[7] K. Ramchandran, A. Ortega and M. Vetterli, "Bit allocation for dependent quantisation with applications to multiresolution and MPEG video coders," *IEEE Transactions on Image Processing*, vol. 3, pp. 533–545, September 1994.