EFFICIENT SCALABLE DCT-BASED VIDEO CODING AT LOW BIT RATES

Michael Gallant and Faouzi Kossentini

Department of Electrical and Computer Engineering,
University of British Columbia,
Vancouver, BC, V6T 1Z4, Canada
E-mail: {mikeg,faouzi}@ece.ubc.ca

ABSTRACT

It is well-known that flexibility and error resilience are significantly improved by employing a scalable bit stream. The major drawback of multi-layered representations within a motion compensated (MC) discrete cosine transform (DCT) based framework is the increase in bit rate as compared to a single-layered representation having the same frequency, spatial and temporal resolution as in the highest layer of the multi-layered representation. Using rate-distortion (RD) optimization techniques, we can improve the compression efficiency of MC-DCT based SNR and spatially scalable video coding framework. We first show how RD optimization techniques can be applied independently for each layer. We then extend the framework to consider coding decisions jointly across layers.

1. INTRODUCTION

In scalable video coding systems, representations are available in a series of relative degrees of resolution. The base layer of video, representing a given resolution or picture quality, is encoded independently of other layers while the subsequent layers of video, representing increased resolution or enhanced picture quality, are encoded dependently, with each following layer coded with respect to the previous layers. This provides additional flexibility in the sense that the scalable bit stream can be manipulated at any point after it has been generated. The capability is desirable in order to counter specific limitations and differences, including constraints on bit rate, decoder complexity, channel error characteristics and display resolution that, in the case of multipoint and broadcast video applications, cannot be foreseen at the time of encoding. Typically, a layer represents a change of scale in frequency, spatial, or temporal resolution. An SNR enhancement layer attempts to recover the coding loss between the reconstructed reference layer picture and the original picture. A spatial enhancement layer attempts to recover the coding loss between an upsampled version of the reconstructed reference layer picture and a higher resolution version of the original picture.

The major drawback of multi-layered representations within an MC-DCT framework is the increase in bit rate as compared to a single-layered representation having same frequency, spatial and temporal resolution as in the highest layered of the multi-layered representation. This increase in bit rate is due to side information overhead, variable-length coding inefficiencies, and the differing statistics of the error signal. Consequently, much of the research in the area of scalability has focused on non MC-DCT based techniques having inherently scalable properties, e.g. subband techniques. Unfortunately, these techniques generally suffer from inferior compression efficiency due to the difficulty of effectively including motion within subband schemes. Furthermore, the ubiquity of MC-DCT based technology and the inclusion of syntax extensions to support scalable coding within newer MC-DCT based video coding standards [1] suggest that scalability be addressed within the MC-DCT framework. We employ well-known RD optimization techniques to improve compression efficiency, based on Lagrangian minimization [2]

$$J = D + \lambda R. \tag{1}$$

We choose the Lagrangian rate-distortion functional as it provides an elegant framework for determining the optimal choice of motion vectors and prediction modes by weighting a distortion term against a resulting rate term for a particular choice of coding parameters. Here, D is defined as some distortion measure, typically the sum of absolute error (SAE) or sum of squared error (SSE). For motion estimation, R, is defined as the sum of the rates for the vertical and horizontal macroblock (or block) motion vector candidates. For mode decision, R is defined as the sum of the rates to encode the target macroblock, including all control, motion, and texture information. The Lagrangian multiplier λ is the weighting parameter that governs the rate-distortion tradeoffs. By considering the various possible combinations of permissi-

This work was supported by the Natural Sciences and Engineering Research Council of Canada, Rogers Communications Inc and the Association of Universities and Colleges of Canada.

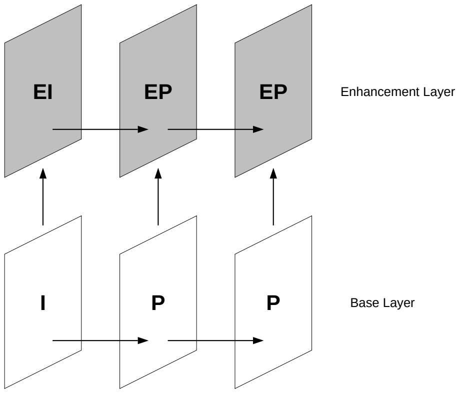

Figure 1: Illustration of possible prediction modes for enhancement layers.

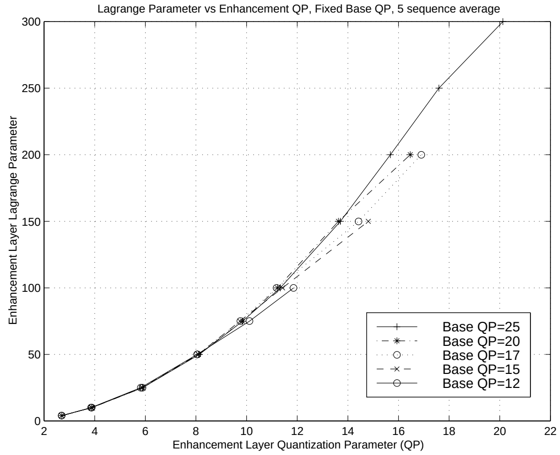

Figure 2: Relationship between enhancement layer Lagrangian and quantization parameters for SNR scalability.

ble coding parameters, we can select the set that produces the minimum Lagrangian cost for a particular value of λ.

A good review of RD optimized techniques for motion estimation and coding mode decisions is available in [3]. Briefly, in MC-DCT based video coding systems, RD optimized motion estimation selects the motion vector that minimizes the Lagrangian cost between the target macroblock (or block) and the macroblock (or block) in the reference picture displaced by the candidate motion vector. RD optimized mode decision selects the coding mode among the

- FORWARD-SKIPPED,
- FORWARD-INTER,
- FORWARD-INTER4V and
- INTRA

modes that minimizes the Lagrangian cost. Here INTER4V refers to the use of four motion vectors for each 16×16 pixel macroblock. Treating motion estimation and mode decision independently and considering each coding unit (macroblock) independently leads to a locally optimal decision for the given λ and coding unit.

2. RATE-DISTORTION OPTIMIZATION FOR SCALABLE CODING

Extending our work on RD optimized H.263 coding [1] from the single layered [4] to the multi-layered framework, we incorporate the additional inter-layer coding dependencies present in a multi-layered framework into the set of permissible coding parameters. Figure 1 illustrates how enhancement layer pictures can have macroblocks (or blocks) forward predicted from a temporally previous enhancement layer picture or upward predicted from a temporally simultaneous reference layer picture. Thus, for RD optimized mode decision in the enhancement layer of an H.263 layered coder, the possible coding modes that we consider are

- FORWARD-SKIPPED,
- FORWARD-INTER,
- FORWARD-INTER4V,
- UPWARD-INTER,
- BI-DIRECTIONAL-INTER,
- BI-DIRECTIONAL-INTER4V and
- INTRA.

Here UPWARD refers to prediction from the macroblock at the same spatial location in the temporally simultaneous reference layer picture (with an assumed motion vector of $(0, 0)$), and BI-DIRECTIONAL refers to prediction formed from the average of the UPWARD and FORWARD predictors.

2.1. Choice of Lagrangian Parameter

To eliminate the time-consuming task of calculating a suitable value of the Lagrangian parameter λ for each frame, we attempt to model the choice of λ as a function of the reference and enhancement layer quantization parameters, Q_{base} and $Q_{enhance}$ [1]. This allows the RD optimized framework to work easily in conjunction with rate control techniques that control the average bit rate by adjusting the quantization parameters. RD optimized mode decision in the enhancement layer then selects the coding mode among the seven possible enhancement layer modes described in the previous section.

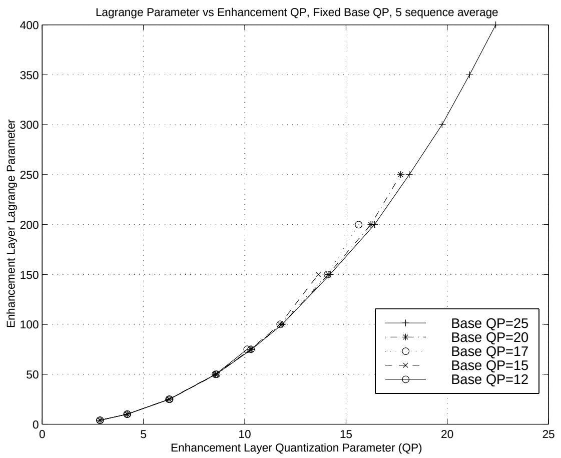

Figure 3: Relationship between enhancement layer Lagrangian and quantization parameters for spatial scalability.

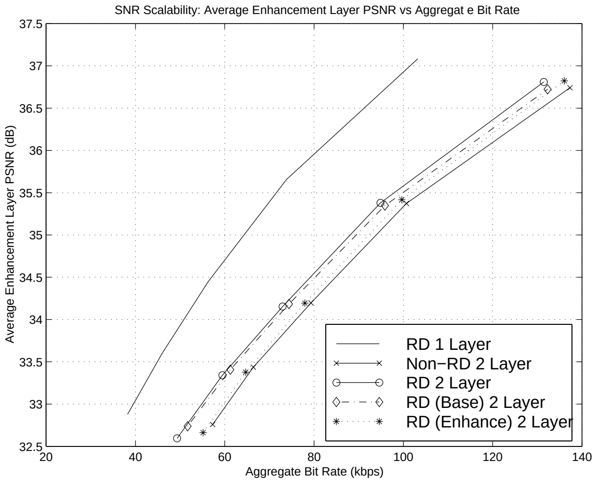

Figure 4: SNR Scalability: Rate distortion optimization in individual layers and in both layers. Average enhancement layer PSNR vs aggregate bit rate, FOREMAN, QCIF, 10fps.

In Figure 2, we plot the average SNR enhancement layer quantization parameter $Q_{enhance}$ obtained by fixing $\lambda_{enhance}$ and allowing $Q_{enhance}$ to vary. Results were obtained by gathering data for five different sequences, using six different values of Q_{base} for each sequence, and nine different values of $\lambda_{enhance}$ for each value of Q_{base}. For fine enhancement layer quantizers, i.e. less than 10, the relationship between the enhancement layer quantization and Lagrangian parameters is well approximated by the second order polynomial

$$\lambda_{enhance} = 0.8 \times \left(\frac{Q_{enhance}}{2}\right)^2 - 0.26 \times \left(\frac{Q_{enhance}}{2}\right) - 1.23. \tag{2}$$

For coarse enhancement layer quantizers, i.e. greater than 10, the relationship between the enhancement layer quantization and Lagrangian parameters is well approximated by the linear equation

$$\lambda_{enhance} = \alpha \times \left(\frac{Q_{enhance}}{2}\right) - \beta, \tag{3}$$

where

$$\alpha = 0.81 \times \left(\frac{Q_{base}}{2}\right) + 3 \tag{4}$$

and

$$\beta = 9.165 \times \left(\frac{Q_{base}}{2}\right) - 66. \tag{5}$$

In Figure 3, we plot the average enhancement layer quantization parameter obtained from similar experiments conducted for spatial enhancement layers. For fine enhancement layer quantization parameters, i.e. less than 10, the relationship between the enhancement layer quantization and Lagrangian parameters is well approximated by the second order polynomial

$$\lambda_{enhance} = 0.81 \times \left(\frac{Q_{enhance}}{2}\right)^2 - 1.05 \times \left(\frac{Q_{enhance}}{2}\right). \tag{6}$$

For coarse enhancement layer quantization, i.e. greater than 10, the relationship between the enhancement layer quantization and Lagrangian parameters is well approximated by the second order polynomial

$$\lambda_{enhance} = \alpha \times \left(\frac{Q_{enhance}}{2}\right)^2 - \beta \times \left(\frac{Q_{enhance}}{2}\right), \tag{7}$$

where α and β depend on Q_{base}, as determined by plotting the empirical values against Q_{base}, and are given by

$$\alpha = 0.003 \times \left(\frac{Q_{base}}{2}\right)^2 - 0.159 \times \left(\frac{Q_{base}}{2}\right) + 2.780 \tag{8}$$

and

$$\beta = 0.034 \times \left(\frac{Q_{base}}{2}\right)^2 - 1.630 \times \left(\frac{Q_{base}}{2}\right) + 21.378. \tag{9}$$

3. RESULTS

The coder employed for the simulations is based on our public TMN-3.2.0 coder [5]. While the public coder only supports one enhancement layer, our modifications allow us to generate up to fifteen enhancement layers, the maximum permissible by the syntax, However, for clarity we restrict ourselves here to using one enhancement layer.

We incorporate Equations (2) - (9) into our coder and generate two layer bit streams with both the non-RD optimized coder and the RD optimized coder. We also generate RD optimized single layer bit streams with the same resolution as the second layer of the two layer bit streams.

3.1. SNR Scalability

In Figure 4, we illustrate the rate-distortion performance of five coders. Four of the coders produce two layer bit streams and one coder produces a single layer bit stream. The single layer coder uses the same fixed quantization parameter that is used in the enhancement layer by the scalable coders. As expected, none of the scalable coders achieve the rate-distortion performance of the non-scalable coder.

The performance of the non-RD optimized scalable coder is 1.5 - 1.7 dB lower in PSNR than that of the non-scalable, i.e. single layer, coder. If RD optimization is performed in the enhancement layer only, the scalable coder incurs a 1.6 dB decrease in PSNR as compared to the single layer coder. For RD optimization in the base layer only, a 1.4 dB decrease in PSNR (approximately 29 percent increase in bit rate) is observed for the scalable coder. If we employ RD optimization in both the base and enhancement layers, the scalable coder suffers only a 1.2 dB decrease in PSNR. Thus, while we are still somewhat far from matching the performance of a single layer coder, RD optimization of both base and enhancement layers improves the rate-distortion performance of scalable coding by as much as 0.5 dB.

Of interest is the observation that RD optimization in the base layer alone provides more gains, in terms of rate-distortion performance, than RD optimization in the enhancement layer alone. One might conclude that this is due to the proportion of the total bit rate taken by the base layer being greater than that taken by the enhancement layer. However, further experiments revealed that this is mainly due to RD optimization in the base layer significantly reducing the amount of intra-coded macroblocks, which are the most expensive in terms of bits. On the other hand, in the enhancement layer, although the intra-mode is a possible coding mode, it is rarely used by even the non-RD optimized coder. This basically eliminates the potential for RD optimization in the enhancement layer to produce the significant savings obtainable by non-intra coding of the macroblocks.

3.2. Spatial Scalability

In Figure 5, we illustrate the rate-distortion performance of six coders. Four of the coders produce two layer bit streams and two produce single layer bit streams. The same fixed quantization parameter is employed in both the base and enhancement layers of the layered coders. The base layers have QCIF resolution while the enhancement layers have CIF resolution. The single layer coders also use the same fixed quantization parameters, and code the same resolution as the enhancement layer of the two layer coders, i.e. CIF.

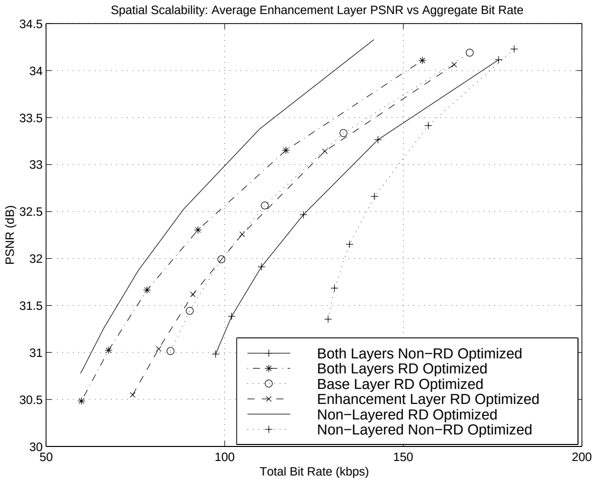

Figure 5: Spatial Scalability: Rate distortion optimization in individual layers and in both layers. Average enhancement layer PSNR vs aggregate bit rate, FOREMAN, QCIF/CIF, 10fps.

First, we look at the performance of the single layer coders relative to the layered coders. Notably, the non-RD optimized single layered coder is outperformed by all layered coders. As FOREMAN contains high motion, camera motion, and occlusions, a significant proportion of P-picture macroblocks are intra-coded in the non-layered coder, for CIF resolution pictures. In the layered coder, most of this intra-coding is performed at the base layer, for QCIF resolution pictures. Therefore, blocks that are intra-coded by the single layer coder are, in the enhancement layer pictures of the layered coder, predicted from the upsampled base layer pictures. As expected, none of the layered coders achieve the rate-distortion performance of the RD optimized single layer coder as RD optimization in the single layer coder can significantly reduce the number of macroblocks that are coded as intra.

Next we look at the performance of the different layered coders relative to the single layer RD optimized coder. The non-RD optimized layered coder incurs a 1.1 - 1.9 dB decrease in PSNR. If RD optimization is performed in the enhancement layer only, the layered coder incurs a 0.8 - 1.35 dB decrease in PSNR. For RD optimization in the base layer only, a 0.75 - 1.4 dB decrease in PSNR is observed for the layered coder. If we employ RD optimization in both the base and enhancement layers, the layered coder suffers only a 0.3 - 0.5 dB decrease in PSNR. Thus, while we still cannot match the performance of an RD optimized single layer coder, we observe that RD optimization of both base and enhancement layers improves the rate-distortion performance

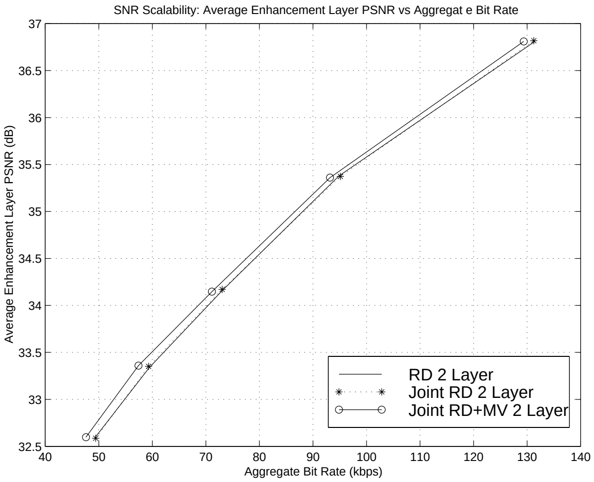

Figure 6: SNR Scalability: Joint rate-distortion optimization among layers and reuse of motion vector field. Average enhancement layer PSNR vs aggregate bit rate, FOREMAN, QCIF, 10fps.

of scalable coding by as much as 1.4 dB. We also observe that RD optimization in the base layer alone provides similar gains, in terms of rate-distortion performance, as RD optimization in the enhancement layer alone. This is because, while RD optimization in the base layer significantly reduces the amount of intra-coded macroblocks, which are the most expensive in terms of bits, RD optimization in the enhancement layer operates on pictures having higher spatial resolution. This results in good improvements in coding efficiency for both the base and enhancement layers.

3.3. Joint Optimization

We have observed that the overall improvement in PSNR is not simply the sum of the improvements in the individual layers. Rather, the rate-distortion improvements achieved in the base layer limit somewhat the gains achievable by RD optimization in the enhancement layer. This suggests further gains can be achieved by considering coding mode decisions for the base and enhancement layers jointly.

Furthermore, we can obtain additional gains by reusing the motion vector field for all layers having the same spatial resolution. We base our motion vector selection on the enhancement layer images, as sub-optimal motion vector choices can be better absorbed by coarser quantization in the base layer.

In Figure 6, we illustrate the rate-distortion performance of three coders. The first is again our RD optimized layered coder. The second coder employs a joint optimization whereby the coding modes for the base and enhancement layer macroblocks are selected to minimize the cost function

$$J_{total} = J_{base} + J_{enhance}. \qquad (10)$$

where the component costs are computed as in Equation 10. The third coder also employs joint optimization as well as motion vector field reuse, as outlined above. Clearly, joint optimization provides little improvement over independently making coding mode decisions within each layer. The improvement in PSNR is at most 0.1 dB. Reuse of the motion vector field provides an additional 0.2 dB.

4. CONCLUSION

We have presented a simple relationship governing the choice of $\lambda_{enhance}$ for SNR and spatially scalable MC-DCT based video coding. Using this relationship, we extend our RD optimized coder to incorporate scalable coding. In the case of SNR scalability, for the two layer bit streams, we obtain a 0.5 dB improvement in PSNR by using RD optimization in both the base and enhancement layers. Employing joint optimization and reusing the motion vector field increases this improvement to 0.7 dB. In the case of spatial scalability, for the two layer bit streams, we obtain a 0.6 - 1.4 dB improvement in PSNR by using RD optimization in both the base and enhancement layers.

5. REFERENCES

[1] ITU Telecom. Standardization Sector of ITU, "Video Coding for Low Bitrate Communication," *ITU-T Recommendation H.263 Version 2*, January 1998.

[2] H. Everett III, "Generalized lagrange multiplier method for solving problems of optimum allocation of resources," *Operation Research*, vol. 11, pp. 399–417, 1963.

[3] G.J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE Signal Proc. Magazine*, pp. 74–90, Nov. 1998.

[4] M. Gallant, G. Cote, and F. Kossentini, "Description of and results for rate-distortion optimized coder," in *Q15-D-49, ITU-T Q15/SG16*, Tampere, Finland, Apr. 1998.

[5] Signal Processing and Multimedia Laboratory, University of British Columbia, "TMN 10 (H.263+) Encoder/Decoder, Version 3.2.0," *TMN 10 (H.263+) codec*, Sept. 1998.