PERFORMANCE ENHANCEMENT OF H.264 CODEC BY LAYERED CODING

Hoda Roodaki[(1)], *Hamid R. Rabiee*[(1)], *Mohammad Ghanbari*[(1,2)],
(1) AICTC Research Center, Sharif University of Technology, Tehran, Iran
(2) University of Essex, UK

ABSTRACT

Transmission of video over error prone and still bandwidth limited wireless channels demand high compression efficiency and resilience to packet losses and errors. Scalable or layered video coding applied to highly compression efficient codecs is an ideal solution to the problem. However, scalability reduces compression efficiency of the coders. In this paper we show how compression efficiency of two-layer SNR scalable video coders can be retained via joint base-enhancement layer optimization. Simulation results show that joint base-enhancement layer optimization significantly outperforms separate optimization of the layers, and it closely follows the compression performance of the single-layer optimized codec.

***Index Terms*—**Video codecs**,** Optimization methods

1. INTRODUCTION

Digital video compression techniques have played a key role in recent multimedia communications. The limitation of bandwidth in communication channels and storage media demands more efficient video coding methods. On the other hand, introducing new applications and advances in multimedia technology demands video coding methods to include more complex and advanced features. Therefore, in 2001 VCEG and MPEG formed a Joint Video Team (JVT) to finalize the project called H.264/AVC or MPEG4 part 10. H.264/AVC has achieved substantial superiority of video quality over the previous standards. This means that H.264 offers significantly higher quality levels with the same bit rates [1]. Rate-distortion optimization used in this codec has significantly improved its compression efficiency [2], [3]. In addition, to adapting H.264 to applications involving bit errors and packet losses, a number of error resilience techniques are provided in the standard. These techniques make the generated bitstream robust to the bit errors and packet losses [4].

Scalability or layering technique is considered as a powerful method to increase resiliency of video codecs to channel errors and losses. In this method a video bitstream is partitioned into layers so that the base layer is independently decodable, and decoding of the enhancement layers, will be depended on the decoding of their lower layers. In the SNR scalable coder, the decoded base layer produces a low quality (SNR) video and the enhancement layers add the additional data necessary to improve video quality further. If the base layer is well protected against the channel error, a minimum picture quality (quality of the base layer) without drift would be guarantied. Layering enables the system to apply different error protection for different layers (or packets) for transmission. The high-priority packets could be transmitted with better error protection and the low-priority ones with less or no error protection or even omitted [5].

In recent years several methods have been proposed for bringing scalability into the H.264 standard codec. In [6], every inter-coded block in the enhancement layer is firstly predicted by one of the upward, direct or forward prediction modes. In the upward mode no additional MV is sent for the block and the reference picture is the base layer reconstructed picture. This mode is especially useful when the base layer MB is finely coded and there is a small residual data to be coded in the enhancement layer. Furthermore, this mode prevents error propagation and picture drift in case data of the enhancement layer is corrupted.

In [7], we have another technique for layered video coding. In this method, at the encoder side the reference picture used for motion compensation must also be the addition of the predicted signal from both layers. One can easily find that feedback information from the second layer's inverse quantization is added to the base layer. The purpose of this design is to increase the accuracy of motion estimation. However, error occurring in the second layer generally propagates to the base layer.

The Joint Video Team (JVT) of the ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG) also proposed the scalable extension of H.264/MPEG4-AVC. The basic concepts for providing temporal, spatial and SNR scalability are analyzed in this scalable video coding design [8]. For SNR scalability, coarse-gain scalability and fine-gain scalability are distinguished.

However the main disadvantage of the scalable coders is a significant increase in its bit rate compared to the single-layer coder with the same quantization parameter [9]. Reasons for the extra overhead bits over the single layer are well analyzed in [9].

This fact restricts the applicability of these scalable coders to situations where the quality of the service is critical. There are also several methods to optimizing a SNR scalable coder. In [10] it has been tried to optimize the second layer bit rate by considering the zero coefficients at this layer. In this method for a quantization parameter q, the distortion and the allocated bit for an arbitrary group of DCT blocks D and R, respectively, the nonzero coefficients are selectively dropped in order to meet either a lower target distortion, $\hat{D}$ or a lower bit rate , $\hat{R}$, within the bounds $\hat{D} < D$ or $\hat{R} < R$; noting that if $\hat{D} = D$ or $\hat{R} = R$ none of the nonzero coefficients are dropped. As another method, instead of either keeping or dropping coefficients, we may adjust them by an integer amount; in this way a much finer control over rate and distortion is achieved [10].

In this paper, we propose a new method for optimizing a SNR scalable coder. This method can be used by any H.264 as well as other standard SNR scalable coders, since the goal of the method is optimizing a SNR scalable coder and it is independent of the scalable method which is used by the coders.

In this joint optimization method the information of all layers, base and enhancements, are optimized jointly. For simplicity of derivations, we have confined our analysis on only two-layer SNR scalable coder, though the same principle can be extended to more layers and other scalability methods. It is argued that for the temporal and spatial scalability, where the interlayer coding distortion is loosely related, separate optimization of the layers can be used. In contrast, for SNR scalability, where the compression distortions between the layers are strongly related to each other, joint optimization of the layers, is beneficial. So we chose SNR scalable coder for our investigations.

The reminder of this paper is organized as follows. Section 2 presents detailed description of a two-layer SNR scalable optimizer. Experimental results are provided in Section 3, followed by concluding remarks in Section 4.

2. PROPOSED ALGORITHM

A simple method for the optimization of a SNR scalable coder is to optimize each layer independent of the other layers. Here the conventional Lagrangian optimization is applied such that, for the given target bit rate of a layer, its Lagrangian cost function is minimized [11]. We call this method Separate-optimization and the Lagrangian function of this separate optimization for a two-layer Scalable coder can be formulated as:

$$\begin{Bmatrix} J_b = D_b + \lambda_b R_b \cdots R_b \le R_{btg} \\ J_e = D_e + \lambda_e R_e \cdots R_e \le R_{etg} \end{Bmatrix} \qquad (1)$$

Thus for a two-layer encoder with the target bit rates of R_{btg} and R_{etg} at the base and enhancement layers, the quantization parameters of their quantization parameters, Q_b and Q_e, are chosen such that their respected cost functions J_b and J_e are independently minimized. In [2] the Lagrangian parameter was recommended to be related to the quantization parameter. Hence according to [2] for a two-layer coder, Lagrangian parameters λ_b and λ_e of the base and enhancement layers may be written:

$$\begin{Bmatrix} \lambda_b = 0.85 \times 2^{\frac{q_b - 12}{3}} \\ \lambda_e = 0.85 \times 2^{\frac{q_e - 12}{3}} \end{Bmatrix} \qquad (2)$$

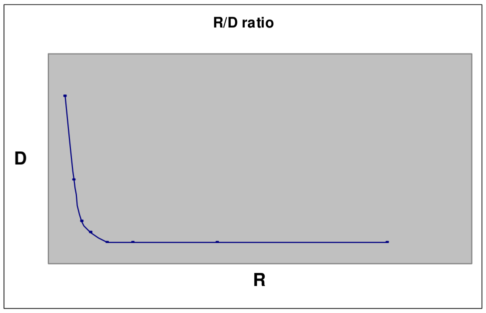
The quantization parameter of each layer and hence their corresponding Lagrangian parameters are varied such that after coding of a number of macroblocks (e.g. within a video slice), equation (1) is satisfied.

It can be argued that a better optimization can be achieved if layers are jointly optimized. The main reason is that layers are not independent of each other, and in particular for SNR scalable coder that the enhancement layer is the residual coding distortion of the base-layer, any optimization on the base layer, that affects its quantization distortion, will be the input to be coded at the enhancement layer. In some implementations of the codecs, such as the one we have proposed for H.264, the enhancement layer can use predictions from both base and enhancement layers [6]. Undoubtedly enhancement layer predictions from the base layer, make inter-layer dependency even stronger and a need for joint-optimization.

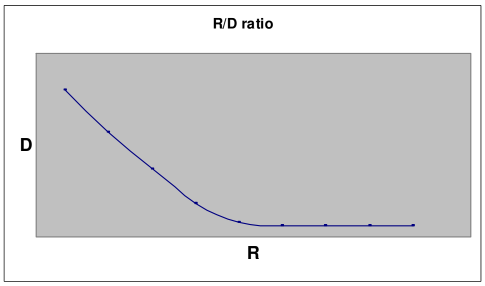
Another reason that necessitates joint optimization of the layers is the fact that layers may have different rate-distortion (R-D) characteristics. For example, if the R-D functions of the base and enhancement layers of a two-layer SNR scalable coder follow those of Figures 1a and 1b, then due to sharper decay of distortion in Figure1a than that of Figure 1b, use of any incremental bits in Figure 1a, will lead to a larger distortion decay than Figure 1b. This is very much in line with the optimization principles that the Lagrangian optimizer tries to achieve.

Thus in optimization of rate and distortion we have more freedom to spend the additional bits either at the base layer or use it at the enhancement layer, of course provided the bit rate budget of the base layer is still in the limit, but for the enhancement layer, now it is the total bit rate that should not be exceeded. Thus a two-layer joint-optimization can be written as:

$$\begin{Bmatrix} J_b = D_b + \lambda_b R_b \cdots R_b \le R_{btg} \\ J_e = D_e + \lambda_b R_b + \lambda_e R_e \cdots R_e + R_b \le R_{total} \end{Bmatrix} \qquad (3)$$



(a)



(b)

Figure 1. Rate-Distortion function (a) sharper distortion decay and (b) smoother rate of decay

The main difference between these two separate and joint optimizations is that, while in the separate optimization, the final distortion (which is that of the enhancement layer) only changes with the enhancement layer bit rate variations and is under the influence of only λ_e,, that is

$$-\frac{\partial D_e}{\partial D} = \lambda_e \frac{\partial R_e}{\partial D} \tag{4}$$

But for the joint optimization, the rate of decay of the final distortion, D_e, is given by:

$$-\frac{\partial D_e}{\partial D} = \lambda_b \frac{\partial R_b}{\partial D} + \lambda_e \frac{\partial R_e}{\partial D} \tag{5}$$

That is, the final distortion not only changes with the bit rate changes of either of the layers, but also with different weighting factors of rate changes (i.e. λ_e and λ_b). This was in fact reflected in the two R-D curves given in Figure 1.

This two-layer optimization may be implemented as follows. Assume that optimization is carried out within one video slice, then if the initial quantizer parameters of the base and enhancement layers are not known, we might use the separate-optimization method of equ (1) to derive the initial quantization step sizes of Q_b and Q_e. Now the base layer is optimized on its own criterion and for the optimization of the enhancement layer:

1) Increment the base-layer quantizer parameter, Q_b, by one, which will reduce its bit rate accordingly. Change the quantization parameter of the enhancement layer, Qe, such that the cost function of the joint-Lagrangian optimizer of equ (3) is minimized. In this case, we expect the quantizer parameter of the enhancement layer to be reduced, such that the excess bit rate is compensated by the reduced bit rate of the base layer.
2) Decrement the base layer quantization parameter by one and similar to step (1) carry out the joint optimization equ (3), which will result in a larger enhancement layer quantizer parameter.
3) Chose the minimum of the two cost functions in the above two steps. This will determine the final base and enhancement layer quantization parameters and their bit rates.

Note that in the above optimization, we still use the Lagrangian parameters λ_b and λ_e of separate optimization derived from equ (2), which may not be ideal. Note also that relation between the Lagrangian parameter λ and the quantization parameter Q in equ (2) although has been defined imperially [2], but they are in fact based on the assumption that distortion D and the energy are logarithmically related to each other. This may be true for the base layer, which has the identical characteristics of a single layer encoder, applied to video, but is not true for the enhancement layer, which codes the residual error. In the enhancement layer the resulting quantization distortion is more uniform than that of a single or base layer data. Thus we need more realistic Lagrangian parameters that should be independent of a model and are better matched to the actual data, to be used in the above 3-step joint optimization. This is best done, by calculating the slopes of the R-D functions of each layer.

This is done in the following manner. From the data of the above 3-step joint-optimization, we calculate the slopes of the curves as:

$$\lambda_b = \frac{D_{b2} - D_{b1}}{R_{b1} - R_{b2}} \tag{6}$$

$$\lambda_e = \frac{D_{e2} - D_{e1}}{R_{e1} - R_{e2}}$$

Where D_{b2} is the distortion of base layer at the incremented quantizer parameter of Q_b+1 and the bit rate of R_{b2}. D_{b1} is the distortion of base layer at the decremented quantizer parameter of Q_b -1 and the bit rate of R_{b1}. These values for the enhancement layer are respectively D_{e2}, R_{e2}, D_{e1} and R_{e1}.

We now go back to the 3-step joint optimizer and replace the new Lagrangian parameters with the old ones. The final coding distortion, D_e, and the total bit rate $R_{total}=R_b + R_e$ might slightly change.

Note also that in the joint optimization we might increment or decrement the base layer quantizer step size by more than one step. This is of course picture content dependent, and since larger deviations will make base-layer sub-optimum, then joint optimization has to be carried out with great care. In fact in order that equ (3) to be truly representative of the slopes of the R-D curves, bit rate deviations should be small, that make the Lagrangian parameters to be close to the tangents to the curves at the optimum quantization parameter Q.

3. EXPERIMENTAL RESULTS

In the experiments, the x264 test platform was adopted instead of H.264/AVC JM reference software, as it is one of the H.264 open source encoders [12]. Its development started in 2004, and it has been used in many popular applications like *ffdshow*, *ffmpeg* and *MEncoder*. In a recent study, x264 proved to have a better quality than several commercially available products of H.264/AVC encoders [13]. Throughout the experiments we have used standard test image sequences of QCIF sizes and the allocated bit rate to the base layer was made equal to the enhancement layer, i.e. R_b=R_e. Figure 2 shows the average PSNR of “Foreman" and “Carphone” for Joint and separate optimization methods. As a bench mark, the PSNR of optimized Single-layer is also shown.

4. CONCLUSION

In this paper we have proposed a new method for optimizing a SNR scalable coder. In the proposed method, instead of using Lagrangian optimizer for each layer separately, both layers are considered simultaneously. The base layer is optimized to its target bit rate, but for the second layer the bit rate of the base-layer is deviated from its optimum position such that the total bit rate and final quality are optimized. Through a regression method, the Lagrangian parameter of each layer is found accordingly.

Through one slice optimization area, the joint-optimization could easily follow its given target bit rate, as smoothly as a single layer or a separate optimized encoder does. Moreover, in terms of quality, the joint-optimizer outperforms the separate optimizer.

5. ACKNOWLEDGMENT

The work described in this paper was supported by a grant from Iran Telecommunication Research Center (ITRC).

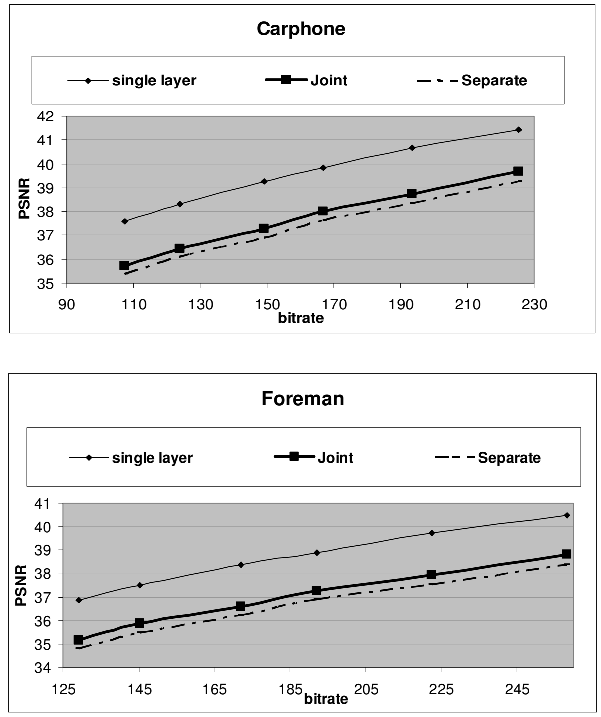


Fig. 2 PSNR of Carphone and Foreman at various bit rates

6. REFERENCES

[1] D. Marpe, T. Wiegand, G. J. Sullivan, “The H.264/MPEG4 Advanced Video Coding Standard and its Applications”, IEEE Communications, 44(8):134-143, Aug. 2006.

[2] T. Wiegand, H. Schwarz, F. Joch, A. Kossentini, and G. J. Sullivan, Rate-constrained coder control and comparison of video coding standards," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, pp. 688{703, July 2003.

[3] M. Ghandi and M. Ghanbari,"A lagrangian optimized rate control algorithm for the H.264/AVC encoder," International. Conferences of Image Processin, ICIP, vol 1, pp. 123-126, October 2004.

[4] M. Ghanbari, ‘Two-layer coding of VBR signals for VBR networks, ‘IEEE Journals of selected Areas of Communications, Vol 7, No. 5, June 1989, pp. 771-781.

[5]Mohammad Mahdi Ghandi, Layered Video Coding for Wireless Communications, PhD. Dissertation, Department of Electronic Systems Engineering University of Essex, January 2006.

[6] M. Mahdi Ghandi and Mohammed Ghanbari, "Layered H.264 video transmission with hierarchical QAM,” Elsevier J. Visual Commun. Image Representation*,* Special issue on H.264/AVC, vol. 17, no. 2, pp. 451 - 466, April 2006.

[7] Yuh-Feng Hsu, Chien-Hua Hsieh and Yung-Chang Chen, " Embedded SNR scalable MPEG-2 video encoder and its associated error resilience decoding procedures", Elsevier Signal Processing:Image Communication*,* vol. 14 , pp. 397-412, 1999.

[8] H. Schwarz, D. Marpe, and T. Wiegand: Overview of the Scalable H.264/MPEG4-AVC Extension, *Proc.* IEEE International Conference on Image Processing (ICIP 2006), Atlanta, GA, USA, October 2006.

[9] M. Ghanbari, Codec Standards: Image Compression to Advanced Video Coding, 1st edition, IEE, 2003.

[10] D. Wilson and M. Ghanbari,” Optimization of MPEG-2 SNR Scaleable Codecs”, IEEE Transaction on Image *P*roceeding, vol. 8, no. 10, October 1999.

[11] Wiegand, T. Schwarz, H. Joch, A. Kossentini, F. Sullivan, G.J, "Rate-constrained coder control and comparison of video coding standards", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 13, NO. 7, July 2003

[12] x264, Available: http://developers.videolan.org/x264.html

[13] “MPEG-4 AVC/H.264 video codec comparison,” CS MSU Graphics and Media Lab, December 2005.