A RATE-DISTORTION OPTIMAL HYBRID SCALABLE/MULTIPLE-DESCRIPTION VIDEO CODEC

Lisimachos P. Kondi

Dept. of Electrical Engineering
State University of New York at Buffalo
Buffalo, NY 14260 USA

ABSTRACT

In this paper, we present a rate-distortion optimal hybrid scalable/ multiple-description video codec. Traditional scalable codecs produce bitstreams which can be partitioned into layers that form a hierarchy. Thus, in order for a particular layer to be useful to the decoder, all hierarchically higher layers also need to be available. Conversely, traditional multiple description codecs produce layers with no hierarchy. Thus, any of the layers can be decoded by itself and produce a video sequence of a certain quality. The more layers are available to the decoder, the better the corresponding video quality. The drawback of multiple description coding is that layers need to be correlated at the expense of compression efficiency. We propose a hybrid scalable/multiple description codec which produces a base layer and two multiple description enhancement layers. The base layer is required for decoding. If one or two of the multiple description enhancement layers are also received, the SNR of the received video sequence is improved. There is no hierarchy in the multiple description enhancement layers. The layers are constructed using a rate-distortion optimal partitioning of the DCT coefficients. Experimental results are presented and conclusions are drawn.

1. INTRODUCTION

There are two main paradigms for layered video coding: Scalable Coding (SC) and Multiple Description Coding (MDC). A scalable codec produces a bitstream that can be partitioned into layers that form a hierarchy. One of the layers is called the base layer and is required for video reconstruction. The other layers, called enhancement layers, can be decoded along with the base layer and produce a video sequence of improved quality. However, in order for an enhancement layer to be useful in decoding, the base layer and all hierarchically higher enhancement layers also need to be available to the decoder. Thus, if a scalable encoder produces one base layer and two enhancement layers and the decoder receives only the base layer and the second enhancement layer, the latter cannot be used in the decoding and the decoder is only able to reconstruct a video sequence of base layer quality.

In contrast with scalable codecs, multiple description codecs produce layers that do not form a hierarchy. Thus, any of the layers can be decoded independently and produce a video sequence of a certain quality. Furthermore, the more layers are available to the decoder, the better the reconstructed video quality. As there is no hierarchy, any layer that is received by the decoder is used in the reconstruction. In order for this to be possible, the multiple description layers need to share some information, thus the layers are correlated. This correlation causes a decrease in coding efficiency.

We propose a Hybrid Scalable/ Multiple-Description Codec (HSMDC) that combines the advantages of the SC and MDC paradigms. The HSMDC codec produces a bitstream that consists of a base layer and several multiple description enhancement layers. The base layer is required for video reconstruction. If one or more multiple-description enhancement layers are received in addition to the base layer, they can be used in the decoding and improve video quality. Thus, there is no hierarchy in the enhancement layers and any received enhancement layer is useful in decoding as long as the base layer has been successfully received. The compression efficiency of HSMDC is better than that of MDC, since only the enhancement layers need to be correlated. Furthermore, as we show in the experimental results, the compression efficiency of HSMDC is close to that of SC, since the correlation between enhancement layers doesn't need to be as high as in the case of MDC. This is due to the fact that the most important information is transmitted with the base layer, which needs not be correlated with any other layer. Thus, HSMDC relaxes the hierarchy of SC by only requiring the base layer to be successfully received and provides non-hierarchical enhancement layers at the expense of a small reduction in compression efficiency.

Scalability is supported by most of the current Motion- Compensated DCT-based (MC-DCT) video compression standards such as MPEG-2, MPEG-4 and H.263. Version 2 of the H.263 standard (also known as H.263+) [1, 2] supports SNR scalability as well as spatial and temporal scalability. SNR scalability implies that the enhancement in quality translates in an increase of the SNR of the reconstructed video sequence, while spatial and temporal scalability imply that the spatial and temporal resolution, respectively, are increased.

The MPEG-4 standard also supports Fine Granularity Scalability (FGS) [3]. In FGS, the video sequence is encoded into a base layer and an enhancement layer. For the enhancement layer, the difference between the original picture and the base layer reconstructed picture is encoded using bit-plane coding of the DCT coefficients. Thus, the enhancement layer bitstream can be truncated at any point while still being able to be decoded (yielding lower video quality).

It has been shown in [4] that, for transmission over error-prone channels, it is advantageous to use scalability and apply stronger error protection to the base layer than to the enhancement layers (Unequal Error Protection). Thus, we can expect a basic reconstructed quality with high probability even during adverse channel conditions. Had we not used scalability but instead protected the whole bitstream equally, there would be a much higher probability of catastrophic errors that would result in a reconstructed video

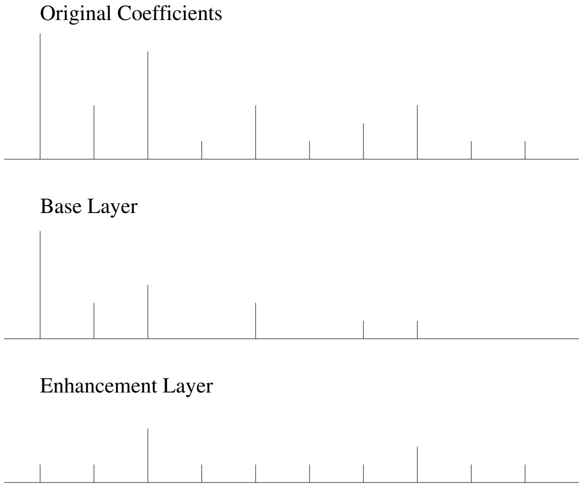

Fig. 1. Proposed partitioning of DCT coefficients for SNR scalability.

sequence of poor quality.

Some of the early theoretical work on multiple description coding appears in [5, 6, 7]. The most recent multiple description image and video coding techniques include, pairwise correlating transforms [8, 9, 10, 11], scalable coding in conjunction with unequal error protection [12, 13], correlating filter banks [14] and coefficient splitting in the DCT domain [15]. Video Redundancy Coding (VRC), which is supported by the H.263 standard [2], can also be seen as a multiple description coding technique.

The proposed Hybrid Scalable/Multiple-Description algorithm operates in two steps. It is first assumed that the DCT of the Displaced Frame Difference (DFD) (or the intensity for intra blocks) is taken and quantized. Then, during the first step, the base layer layer is constructed for each block by subtracting a suitable value from each quantized coefficient (see Fig. 1). The subtracted values then become the enhancement layer. The determination of the subtracted values is optimal in the operational rate-distortion sense. In the second step of the algorithm, the enhancement layer obtained in the first step is converted into two multiple description enhancement layers by selecting a threshold for each block and duplicating into both descriptions the quantized coefficients with values equal to or greater than the threshold while alternating between descriptions the other coefficients, in a fashion similar to the algorithm in [15]. Again, the determination of the threshold is done is optimal in the operational rate-distortion sense. The rest of the paper is organized as follows. In section 2, the algorithm for the determination of the base layer is presented, while in section 3, the determination of the multiple description enhancement layers is discussed. In section 4, experimental results are presented. Finally, in section 5, conclusions are drawn.

2. STEP ONE: DETERMINATION OF THE BASE LAYER

The proposed codec is based on the architecture of the H.263 video compression standard. However, any Motion-Compensated DCT codec can be used as a basis. We assume that the DCT transform of the DFD (or the intensity for intra blocks) is taken and quantized. That is, a triplet (LEVEL,RUN,LAST) is transmitted using suitable VLC tables, where LEVEL is the quantization level of the coefficient, RUN is the number of zero-valued coefficients that precede it and LAST specifies whether the current coefficient is the last in the block. An extra bit is appended to the VLC to denote the sign of LEVEL. Therefore, in the following discussion, LEVEL will refer to the absolute value of the quantization index.

In forming an SNR scalable bitstream the following problem is formulated and solved. Let X be the set of original (unquantized) DCT coefficients in a frame and C the set of quantization levels that results from the quantization of X with quantization parameter QP.

Our goal is, given a set of DCT coefficients X with corresponding quantization levels C and "dequantized" values $\hat{X}$, to find a set of quantization levels $\tilde{C}$ by subtracting a certain value l_i from each coefficient quantization level C_i, so that a bit constraint is satisfied. The value l_i can be different for each coefficient quantization level C_i. The set of "dequantized" values that corresponds to $\tilde{C}$ is $\tilde{X}$. We will call $\tilde{X}$ a *trimmed* version of $\hat{X}$. The set of quantization levels $\tilde{C}$ is transmitted as the base layer (along with motion vectors and overhead information). Then, given a bit budget for the base layer, our problem is to find $\tilde{C}$ as the solution to the constrained problem

$$\min_{\tilde{C}} [D(X, \tilde{X})|C] \text{ subject to } R(\tilde{C}) \leq R_{budget}, \qquad (1)$$

where $D(.,.)$ and $R(.)$ are the distortion and rate functions, respectively, and R_{budget} is the available bit budget for the base layer.

The problem of Eq. (1) can be solved using Lagrangian relaxation. The problem now becomes the minimization of the Lagrangian cost

$$J_1(\lambda) = D(X, \tilde{X}|C) + \lambda R(\tilde{C}), \qquad (2)$$

and the specification of the Lagrange multiplier λ so that the budget constraint is satisfied.

Without lack of generality, in our implementation of the algorithm, we determine a bit budget for the base layer for a Group of Blocks (GOB). This is done because an outside rate control mechanism updates the Quantization Parameter (QP) at the beginning of each GOB and thus determines the total available bit budget for the GOB (for all scalable layers). The bit budget for the base layer is a fixed percentage of the total available bit budget for the GOB. This percentage is determined by the target bit rates for each scalable layer. In H.263 with QCIF-sized frames, one GOB consists of one line of 16×16 macroblocks (11 macroblocks). Each macroblock consists of four luminance and two chrominance 8×8 blocks. Since the encoding of the DCT coefficients is done independently for each block (except for the dc coefficient of intra blocks which is differentially encoded and transmitted with the base layer anyway), $J_1(\lambda)$ is expressed as the sum of individual Lagrangian costs (one for each block) and the minimization is performed individually for each block, using the same λ [16, 17]. Then, if the bit budget for the whole GOB is met for a specific λ, we are guaranteed that the minimization of the individual Lagrangian costs results in an optimal bit allocation across the whole GOB. The λ for which the bit budget is met is found iteratively. A large λ results in a point in the rate-distortion curve with low rate and high distortion. Conversely, a small λ results in a point with high rate and low distortion. Therefore, a simple method, such as bisection, can be used to find the desired λ iteratively. More sophisticated algorithms, such as, the fitting of a Bezier curve [17], can also be used.

The problem now reduces to finding the set of quantization levels $\tilde{C}$ and corresponding trimmed DCT coefficients $\tilde{X}$ for every block that would minimize the Lagrangian cost of the block

$$J_{1,block}(\lambda) = D_{block}(X, \tilde{X}|C) + \lambda R_{block}(\tilde{C}), \tag{3}$$

for a given λ. The admissible candidate set $\tilde{C}$ is constructed as follows. Each non-zero coefficient in the block with quantization level C_i is either dropped completely or a value $l_i < C_i$ is subtracted from it. Although there is a finite number of admissible sets $\tilde{C}$, the minimization of the Lagrangian cost in Eq. (3) using exhaustive search is computationally prohibitive. The problem has however a structure which can be exploited using Dynamic Programming (DP) for its solution. The details of the DP algorithm can be found in [18].

3. STEP TWO: DETERMINATION OF THE MULTIPLE DESCRIPTION ENHANCEMENT LAYERS

We have thus far presented an optimal algorithm for partitioning a set of quantized DCT coefficients into two layers. We are now ready to split the second layer into two multiple description enhancement layers, by also adding an appropriate amount of redundancy.

Let us assume that we have already partitioned the DCT coefficients into two layers and the set of coefficient quantization levels for the second layer is C_{enh}. We now want to partition C_{enh} into two sets of coefficients, namely C_2 and C_3. The coefficients of the base layer C_1 have already been selected during the partitioning of the coefficients into two layers. Let X_{1+2} be the "dequantized" DCT coefficients when layer 1 (base layer) and layer 2 (first multiple description enhancement layer) are utilized. Similarly, let X_{1+3} be the "dequantized" DCT coefficients when layer 1 (base layer) and layer 3 (second multiple description enhancement layer) are utilized. In order to construct X_{1+2} and X_{1+3}, the decoder adds up the corresponding quantization levels before "dequantization", as in the case of the scalable video codec in [18, 19]. The two sets of coefficients C_2 and C_3 are constructed from C_{enh} by determining a threshold T for each 8×8 image block and duplicating on both C_2 and C_3 those coefficient quantization levels that are equal to or greater than the threshold, while alternating between layers the remaining coefficients. Thus, the first coefficient that is below the threshold will go to C_2, the second one to C_3 and so on. Clearly, the smallest the threshold T, the greater the redundancy introduced. Let T_{GOB} be the set of thresholds for a whole Group of Blocks (GOB). Given the algorithm of the creation of the multiple description enhancement layers, it is expected that both multiple description enhancement layers will have similar rate and distortion. Thus, we choose to define the problem of determining the multiple-description enhancement layers as follows: Determine the set of thresholds T_{GOB} such that

$$\min_{T_{GOB}} [D(X, X_{1+2})|C_{enh}] \text{ subject to } R(C_2)+R(C_3) \leq R_{budget,enh}. \tag{4}$$

where $R_{budget,enh} = R(C_{enh}) \cdot \alpha$, where $1 \leq \alpha \leq 2$ determines the added redundancy. $\alpha = 1$ corresponds to $R(C_2) + R(C_3) = R(C_{enh})$ (no added redundancy) while $\alpha = 2$ corresponds to $C_2 = C_3 = C_{enh}$.

Layers Decoded	PSNR (dB)
Layer 1 (32 kbps total)	28.97
Layers 1+2 (48 kbps total)	30.00
Layers 1+3 (48 kbps total)	30.01
Layers 1+2+3 (64 kbps total)	31.43

Table 1. Performance of the HSMDC for the "Foreman" sequence using $\alpha = 1$.

As in Step One, the constrained optimization problem of Eq. (4) can be converted to an unconstrained optimization problem through the use of Lagrangian multipliers. Thus, we have the Lagrangian cost

$$J_2(\mu) = D(X, X_{1+2})|C_{enh}] + \mu R(C_2). \tag{5}$$

As before, J_2 can be broken into a sum of individual Lagrangian costs, one for each block:

$$J_{2,block}(\mu) = D_{block}(X, X_{1+2})|C_{enh}] + \mu R_{block}(C_2). \tag{6}$$

Then, the appropriate μ that will meet the target bit rate can be determined using a method like the bisection method, as in Step One.

The HSMDC decoder works as follows. If only the base layer is available, it is decoded by itself, just like nonscalable H.263. If the base plus one of the two multiple description enhancement layers are available, the decoder adds up the corresponding quantization indices before inverse quantization. If the base plus both multiple description enhancement layers are available, the decoder first processes the multiple description enhancement layer in order to reconstruct the original enhancement layer produced in Step One. The decoder compares the values of the coefficients of the two layers. If the value of a particular coefficient is the same in both layers, this means that this coefficient was duplicated on both layers in Step Two. If the value of a particular coefficient is zero in one layer and nonzero in the other, this means that the cofficient was alternated in Step Two. This way, the decoder is able to recover the enhancement layer produced in Step One. Then, it adds that layer to the base layer and performs "inverse quantization".

4. EXPERIMENTAL RESULTS

We next present experimental using the proposed Hybrid Scalable/Multiple Description Coder. In order to gauge the performance penalty incurred in utilizing multiple description enhancement layers instead of hierarchical enhancement layers, we compare the proposed HSMDC codec with the three-layer scalable codec in [18, 19]. The "Foreman" sequence was used in the experiments. For both codecs, an external rate control was used to maintain the total rate for all layers at 64 kbps. The resulting frame rate was 7 frames per second. For both encoders, the base layer rate was set to 32 kbps. The HSMDC codec is designed to produce equal-rate multiple description enhancement layers, thus, each multiple description enhancement layer was of rate 16 kbps, making the total rate for all layers $32 + 16 + 16 = 48$ kbps. We also set the first enhancement layer rate of the scalable codec to 16 kbps. Tables 1 and 2 show the results of the HSMDC using a α of 1 and 1.2, respectively. Since the most important video information goes to the base layer, the multiple description enhancement layers need not be as correlated as in traditional multiple description coding. An increase in α translates in an increase of the PSNR

Layers Decoded	PSNR (dB)
Layer 1 (32 kbps total)	28.89
Layers 1+2 (48 kbps total)	30.13
Layers 1+3 (48 kbps total)	29.98
Layers 1+2+3 (64 kbps total)	31.31

Table 2. Performance of the HSMDC for the "Foreman" sequence using $\alpha = 1.2$.

Layers Decoded	PSNR (dB)
Layer 1 (32 kbps total)	29.05
Layers 1+2 (48 kbps total)	30.58
Layers 1+2+3 (64 kbps total)	31.46

Table 3. Performance of the scalable codec in [18, 19] for the "Foreman" sequence.

of decoding layers 1+2 and layers 1+3 (since the redundancy is increased) and in a decrease of the PSNR of decoding layers 1+2+3. It should be emphasized that the rate control keeps all rates approximately equal for all choices of α. Table 3 shows the performance of the scalable codec of [18, 19]. It can be seen that the extra functionality of having multiple description enhancement layers comes at the expense of a slight decrease of the PSNR when using all layers in the decoding.

5. CONCLUSION

We have presented a hybrid scalable multiple description video codec. We have shown that the functionality of having multiple description enhancement layers comes at the expense of a slight decrease of the PSNR when using all layers in the decoding. We are currently working on the application of the proposed codec on wireless video transmission.

6. REFERENCES

[1] International Telecommunication Union Recommendation H.263, "Video coding for low bitrate communications," January 1998.

[2] University of British Columbia, "TMN version 3.2," H.263+ Public domain implementation.

[3] W. Li, "Overview of fine granularity scalability in MPEG-4 video standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, pp. 301–317, Mar. 2001.

[4] R. Aravind, M. R. Civanlar, and A. R. Reibman, "Packet loss resilience of MPEG-2 scalable video coding algorithms," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, no. 5, pp. 426–435, Oct. 1996.

[5] J. K. Wolf, A. Wyner, and J. Ziv, "Source coding for multiple descriptions," *Bell System Technical Journal*, vol. 59, pp. 1417–1426, Oct. 1980.

[6] L. Ozarow, "On a source coding problem with two channels and three receivers," *Bell System Technical Journal*, vol. 59, pp. 1909–1921, Dec. 1980.

[7] A. A. El Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Transactions on Information Theory*, vol. IT-28, pp. 851–857, Nov. 1982.

[8] Y. Wang, M. Orchard, V. Vaishampayan, and A. R. Reibman, "Multiple description coding using pairwise correlating transforms," *IEEE Transactions on Image Processing*, vol. 10, pp. 351–366, Mar. 2001.

[9] V. K. Goyal, J. Kovacevic, R. Arean, and M. Vetterli, "Multiple description coding of images," in *Proceedings of the International Conference on Image Processing*, Chicago, IL, Oct. 1998.

[10] A. R. Reibman, H. Jafarkhani, Y. Wang, M. Orchard, and R. Puri, "Multiple description coding for video using motion compensated prediction," in *Proceedings of the International Conference on Image Processing*, Kobe, Japan, Oct. 1999.

[11] V. K. Goyal and J. Kovacevic, "Generalized multiple description coding with correlating transforms," *IEEE Transactions on Information Theory*, vol. 47, pp. 2199–2224, Sept. 2001.

[12] A. E. Mohr, E. A. Riskin, and R. E. Ladner, "Unequal loss protection: Graceful degradation of image quality over packet erasure channels through forward error correction," *IEEE Journal on Selected Areas of Communications*, vol. 18, pp. 819–828, June 2000.

[13] R. Puri, K. Ramchandran, K. W. Lee, and V. Bharghavan, "Forward error correction (FEC) codes based multiple description coding for internet video streaming and multicast," *Signal Processing: Image Communication*, vol. 16, no. 8, pp. 745–762, May 2001.

[14] X. Yang and K. Ramchandran, "Optimal multiple description subband coding," in *Proceedings of the International Conference on Image Processing*, Chicago, IL, Oct. 1998.

[15] A. R. Reibman, H. Jafarkhani, Y. Wang, and M. Orchard, "Multiple description video using rate-distortion splitting," in *Proceedings of the International Conference on Image Processing*, Thessaloniki, Greece, Oct. 2001.

[16] K. Ramchandran, A. Ortega, and M. Vetterli, "Bit allocation for dependent quantization with applications to multi resolution and MPEG video coders," *IEEE Transactions on Image Processing*, vol. 3, no. 5, pp. 533–545, September 1994.

[17] Guido M. Schuster and Aggelos K. Katsaggelos, *Rate-Distortion Based Video Compression, Optimal Video Frame Compression, and Object Boundary Encoding*, Kluwer Academic Publisher, 1997.

[18] L. P. Kondi and A. K. Katsaggelos, "An operational rate-distortion optimal single-pass SNR scalable video coder," *IEEE Transactions on Image Processing*, vol. 10, no. 11, pp. 1613–1620, Nov. 2001.

[19] L. P. Kondi and A. K. Katsaggelos, "An optimal single pass SNR scalable video coder," in *Proceedings of the International Conference on Image Processing*, Kobe, Japan, 1999.