ON VIDEO SNR SCALABILITY

Lisimachos P. Kondi, Faisal Ishtiaq and Aggelos K. Katsaggelos

Northwestern University
Dept. of Electrical and Computer Engineering
2145 Sheridan Road
Evanston, IL 60208
E-Mail: {lkon,faisal,aggk}@ece.nwu.edu

ABSTRACT

In this paper, we compare two SNR scalable video codecs. The first codec (CODEC1) is a three-layer single-pass quantization algorithm based on H.263 and extends the work presented in [1, 2]. The second codec (CODEC2) implements three layer SNR scalability as described in Annex O of the H.263 standard by re-quantizing the DCT coefficients of the encoding error using finer quantizers than those used in the previous layer. By testing the two algorithms at various base and enhancement layer rates on both high and low motion sequences, the advantages and disadvantages of each codec are discussed and suitable applications are suggested for each one.

1. INTRODUCTION

The importance of signal-to-noise ratio (SNR) video compression algorithms has increased in the past few years. This emergence corresponds to the vast increase of products and applications requiring the transmission of digital video streams. These new applications, including video telephony, teleconferencing, video surveillance, public safety, and video-on-demand, require limiting the bandwidth of the compressed bit stream to less than the capacity of the transmission channel. However, the channel capacity is frequently unknown at the time of compression, especially when the stream is to be broadcast to many users over heterogeneous channels. SNR scalable compression allows a single compression to provide bit streams of multiple quality. In this fashion, the transmitted bit rate can match the available channel(s) without requiring multiple encodings.

SNR scalability has been an integral part of MPEG-2 and has recently been adopted into the latest revision of H.263. Previous work relating to SNR scalability has included [3] where a two-layer SNR optimization scheme was presented for MPEG-2. In [4], the merits of scalability are presented with respect to error resilience while H.263 SNR scalability has been presented in [5] with respect to the Human Visual System (HVS).

In this paper we compare two different SNR scalable video codecs. The first one (CODEC1) [1, 2] combines two separate methodologies for dividing the blocks of discrete cosine transform (DCT) coefficients: spectral selection and successive approximation. The flexible combination of these approaches allows each DCT block to be divided into a fixed number of scans while also controlling the size of each scan. Thus, the transmitted stream can contain any subset of scans from the overall compressed version and thereby both the transmitted bit rate and the quality or SNR are allowed to vary. The second codec (CODEC2) is based on Annex O of the H.263 standard.

This paper is organized as follows. In Section 2, CODEC1 based on a hybrid form of spectral selection and successive approximation is detailed. In Section 3, H.263 SNR scalability is discussed. Results using both types of scalability are presented in Section 4. Finally, conclusions are provided in Section 5.

2. CODEC1

2.1. Spectral Selection

The energy compaction property of the DCT dictates that the majority of the signal's energy is found in the low frequency coefficients. Thus, a typical methodology for dividing a DCT block into scalable scans involves sending only the low frequency coefficients in the first scan, also known as the baselayer. This approach is called spectral selection (SS)[6]. In order to rank each two-dimensional coefficient by its frequency content, a zig-zag ordering is used. In terms of this zig-zag ordering, spectral selection involves transmitting coefficients 0 to $L_1 - 1$ in the baselayer, L_1 to $L_2 - 1$ in scan two, and so on until all coefficients are included.

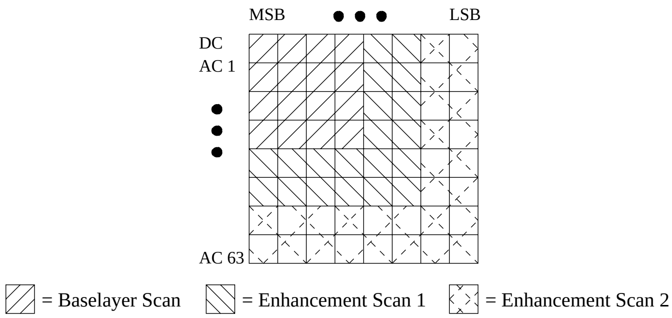


Figure 1: Typical scan definition for dividing an 8 x 8 block of DCT coefficients using both spectral selection and successive approximation

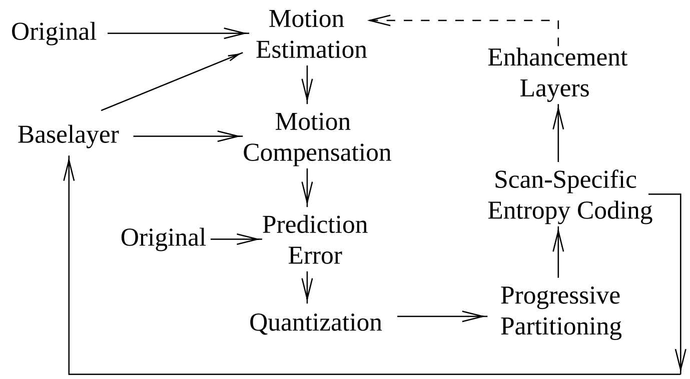


Figure 2: Block diagram of SNR scalable encoder

2.2. Successive Approximation

In contrast to SS, successive approximation (SA) involves including all coefficients in each scan, but increasing the resolution of each coefficient in subsequent scans [7]. This technique corresponds to bit-plane coding.

2.3. Combination of SS and SA

Within a block of DCT coefficients, the low frequency coefficients represent trends, or regions with relatively constant intensity. These coefficients represent the majority of the information content of most image blocks. In contrast to the trends, the high frequency coefficients represent areas of highly varying intensity or edges. While edges are not present in all blocks, the information which they convey is significant to the overall content of the image. Thus, in order to have some tradeoff between edges and trends, a combination of spectral selection and successive approximation can be used to divide a DCT block. An example of a combination of SS and SA is given in Figure 1.

Figure 2 provides a block diagram of the proposed SNR scalable encoder. It only requires a single quantization and a single set of motion vectors. The proposed algorithm uses a block-based motion compensated scheme identical to H.263 [8]. After the DCT of each block is taken, the DCT coefficients are quantized a single time using a fairly small quantizer stepsize. After quantization, we partition the block of DCT coefficients using a combination of spectral selection and successive approximation. In this way, we form a number of scans. Each scan constitutes a subset of the original quantized block of DCT coefficients. Thus, using all scans, we have the complete block of DCT coefficients which were quantized with a small quantizer stepsize. It is important to notice from the block diagram that motion compensation uses only the baselayer from the previous reconstructed frame. This sacrifice is necessary to assure that the decoder can reproduce the encoder's motion compensation without having the enhancement layers.

CODEC1 uses a fixed number of scans of increasing quality within the bit stream. Experiments have shown that the use of three scans gives good results. The scans are defined as follows (8×8 blocks):

- Scan 1: Coefficients 0 to X (all bits except A Least Significant Bits-LSB)
- Scan 2: Coefficients $X + 1$ to 63 (all bits except B Least Significant Bits)
- Scan 3: All remaining bits of coefficients 0 to 63

A rate control algorithm is needed to adjust X, A and B in order to meet the required bit rates for the base layer and the first enhancent layer. We refer to this as the *inside rate controller*. For this paper, we developed a new inside rate controller which extends the rate controller presented in [1, 2]. The inside rate controller adjusts X, A and B at the beginning of each Group of Blocks (GOB). It takes into account the number of bits encoded per layer for the previous GOB and adjusts X, A and B appropriately if the target bit rates were not met.

The overall bit rate (the bit rate of the baselayer plus the two enhancement layers) and the frame rate are controlled using a rate controller similar to the one used in non scalable H.263. We will call this *the outside rate controller.*

CODEC1 utilizes three sets of Variable Length Code (VLC) tables for the transmission of the DCT coefficients. There is a specialized table for each one of the scans.

3. H.263 SNR SCALABILITY (CODEC2)

In Annex O of H.263, SNR scalability is defined as the re-quantization and subsequent re-encoding of the coding error from the previous layer [8]. As enhancement layers are encoded, the SNR quality of the image is improved by the addition of this quantized residual. Given an enhancement layer the encoded base layer frame is subtracted from the original source frame forming a coding error. SNR scalability is described as the encoding of the following signal

$$e_k(\vec{r}) = f_k(\vec{r}) - \hat{f}_k(\vec{r}),$$

where $e_k(\vec{r})$ is the error image given the original source frame $f_k(\vec{r})$ and its base layer reconstructed version $\hat{f}_k(\vec{r})$. This error is then compressed in a fashion similar to encoding the source frame in the base layer using a finer quantizer. This operation is repeated if additional enhancement layers are required.

H.263 defines two separate modes used in encoding the enhancement layer: the EI and EP modes. These modes are equivalent to the I and P modes in the base layer. While EI mode uses only the previous layers' reconstructed frame and the enhancement data, the EP mode uses prediction from the previous enhancement layer, the previous layers' reconstructed frame, and the enhancement data. In our investigations we primarily focus on the EI mode. The codec for an EI mode SNR scalable video codec is shown in Figure 3 noting that due to the predictive nature of H.263, the encoder has a decoder built-in.

In H.263, enhancement layers are designed such that they are syntactically independent of one other. Thus beyond transmitting only residual information for every enhancement layer a host of syntactical data needs to be transmitted. This is in the form of picture headers, start codes, GOB information, and macroblock headers and information. In CODEC1 this overhead information does not need to be transmitted by design and can be a large source of bit savings at very low bitrates.

While the H.263 video coding standard defines the bitstream syntax of a compressed frame, it does not specify the rate control required to achieve a certain quality or compression rate. To regulate the compression such that the target bitrates for each of the layers are achieved while maintaining both good visual and temporal quality, a two-step rate control is used. It is a modified version of the TMN7 rate control [9] which has been further augmented to work with two enhancement layers. Overall, three individual rate controllers operate to regulate the three layers independently. This two-step technique operates by first making the frame drop decisions and deciding on the bit budget for the current frame, followed by the second step working within the frame to provide the macroblocks with the appropriate quantizer step sizes. Since frame drop decisions are no longer an option in the enhancement layers, the rate control for these layers has been modified such that only the bit budget for the current frame is considered.

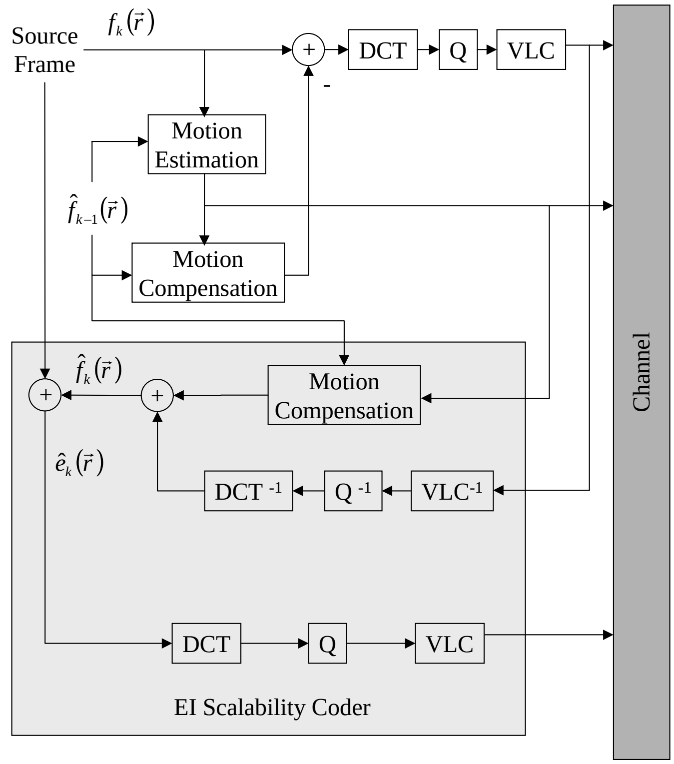


Figure 3: H.263 EI mode SNR scalable video coder

4. EXPERIMENTAL RESULTS

To evaluate the performance of the two scalable techniques, a number of experiments were conducted on two separate video sequences of 300 frames each. The two sequences are the "Foreman" and "Akiyo" sequences used in MPEG-4 testing. The "Foreman" sequence is considered to be a sequence with high motion while the "Akiyo" sequence has comparatively much lower motion. These sequences were tested at two sets of base and enhancement layer bitrates. The set of rates were 14-18-22 Kbps and 28.8-56-128 Kbps. Using this

Bit rate (Kbps)	14	18	22
Foreman CODEC1	27.56	27.93	29.11
Foreman CODEC2	29.00	29.26	29.55
Akiyo CODEC1	34.39	34.66	35.08
Akiyo CODEC2	35.50	35.61	35.74

Table 1: Average PSNR in dB of scalable encoders at 14, 18 and 22 Kbps.

Bitrate (Kbps)	28.8	56	128
Foreman CODEC1	28.72	29.73	34.63
Foreman CODEC2	30.79	32.53	35.56
Akiyo CODEC1	36.73	38.04	41.91
Akiyo CODEC2	38.17	39.06	41.47

Table 2: Average PSNR in dB of scalable encoders at 28.8, 56 and 128 Kbps.

notation, at the 14-18-22 Kbps rates, the base layer is at 14 Kbps and the two SNR enhancement layer rates total at 18 and 22 Kbps, respectively, giving them 4 Kbps ($18 - 14$ Kbps and $22 - 18$ Kbps) each. These are very low bitrates with small layer separations. Results for the two codecs at 14-18-22 Kbps are presented in Table 1. Here we see that for both the "Foreman" and "Akiyo" sequences H.263 SNR scalability (CODEC2) performs better particularly in the base and first enhancement layers. At the highest layer, the performance between the two codecs is comparable.

The second set of rates, 28.8-56-128 Kbps, are rates typical of the progression from modem dial-up rates to ISDN bitrates. It should also be noted that these rates offer much higher layer separations than in the 14-18-22 Kbps case allowing each of the layers with much more bandwidth. This is a more preferred scenario for H.263 SNR scalability because overhead information will not be a significant source of the percentage of bits spent. The performance of the two codecs at these rates are detailed in Table 2. Here we see that for the "Foreman" sequence H.263 SNR scalability (CODEC2) continues to perform better similar to the results in Table 1 but with a larger PSNR differential at the highest layer. While the results for the "Foreman" sequence are consistently better for CODEC2, the "Akiyo" sequence at 28.8-56-128 Kbps shows that CODEC1 has been able to outperform CODEC2 (H.263 SNR Scalability) at the highest layer with the differences for the base and first enhancement layers being less. The encoded bits per frame and PSNR for the "Foreman" sequence are then shown in Figure 4 (a) and (b), respectively, for CODEC1 and in Figure 5 (a) and (b) for CODEC2 at 28.8-56-128 Kbps.

5. CONCLUSIONS

The results in Table 1 show that on a PSNR level, CODEC2 generally outperforms CODEC1. This is also the case in Table 2 for the base and first enhancement layers. For the second enhancement layer, we see that CODEC2 and CODEC1 give similar results. The PSNR difference between the two codecs at the lower layers can be attributed partially to the H.263 codec having better motion prediction and compensation since in the base layer the H.263 codec is simply a non-scalable codec operating at the base layer bitrate. This is in contrast to CODEC2 where motion prediction and compensation are preformed using only a subset of the entire information, providing a lessened quality base layer. However, as we have pointed out earlier, CODEC1 method uses only a single quantization whereas CODEC2 requires three separate quantization steps. In applications where computational complexity must be kept to a minimum, CODEC1 may be more suitable. This is because only a single quantization is necessary and the cost for having three independent bit streams is not incurred. Furthermore, results from the "Akiyo" sequence at 28.8-56-128 Kbps indicate that given a low-motion sequence coupled with low computational complexity, CODEC1 becomes more competitive.

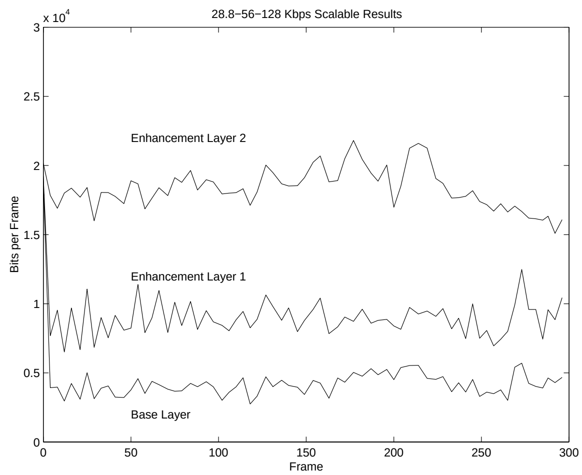
6. REFERENCES

[1] M. A. Robers, "SNR scalable video coder using progressive transmission of DCT coefficients," Master's thesis, Northwestern University, May 1997. Department of Electrical and Computer Engineering.

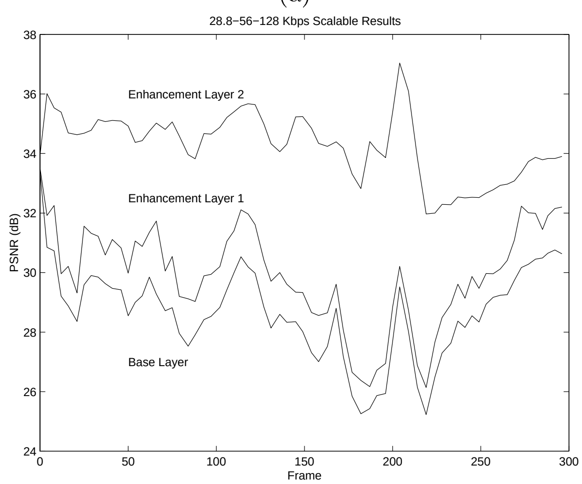
[2] M. A. Robers, L. P. Kondi, and A. K. Katsaggelos, "SNR scalable video coder using progressive transmission of DCT coefficients," in *Proc. SPIE Conf. on Visual Comm. and Image Proc*, pp. 201–212, 1998.

[3] D. Wilson and M. Ghanbari, "Optimization of two-layer SNR scalability for MPEG-2 video," in *Proc. of the Intl. Conf. on Acoustics, Speech and Signal Processing*, vol. 4, pp. 2637–2640, IEEE, 1997.

[4] R. Aravind, M. R. Civanlar, and A. R. Reibman, "Packet loss resilience of MPEG-2 scalable video coding algorithms," *IEEE Trans. on Circuits and*

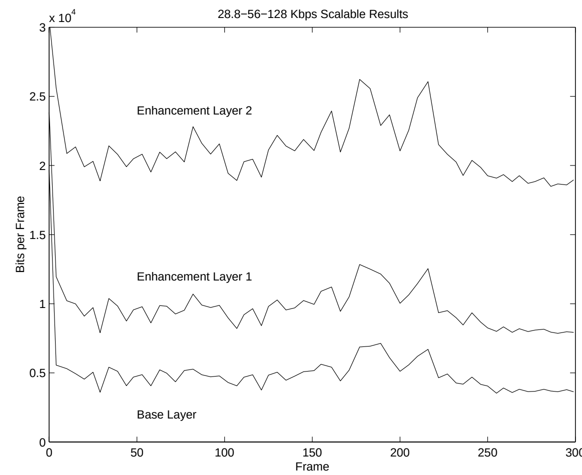


(a)

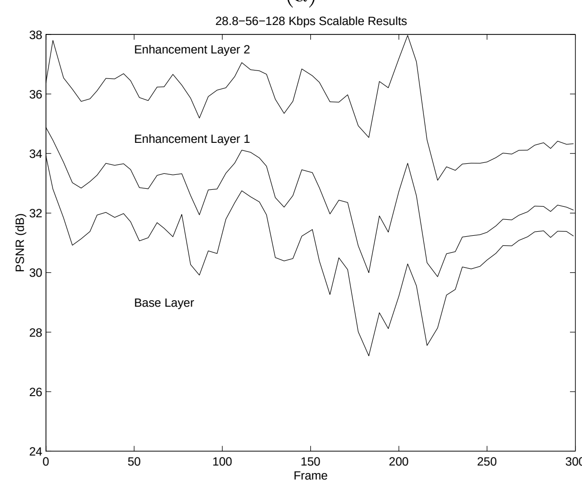


(b)

Figure 4: CODEC1 Bits per frame (a) and PSNR (b) for the "Foreman" sequence at 28.8-56-128 Kbps.



(a)



(b)

Figure 5: CODEC2 Bits per frame (a) and PSNR (b) for the "Foreman" sequence at 28.8-56-128 Kbps.

Systems for Video Technology, vol. 6, pp. 426–435, Oct. 1996.

[5] B. R. Lee, K. K. Park, and J. J. Hwang, "H.263-based SNR scalable video codec," *IEEE Trans. Consumer Electronics*, vol. 43, pp. 614–622, Aug. 1997.

[6] "Information technology- digital compression and coding of continuous-tone still images (JPEG); Recommendation T.84 ." International Organization for Standardization, November 1994.

[7] V. Bhaskaran and K. Konstantinides, *Image and video compression standards: Algorithms and architectures.* Kluwer academic publishers, 1995.

[8] "Video coding for low bitrate communication; DRAFT ITU-T Recommendation H.263." International Telecommunication Union, May 1996.

[9] "Video codec test model TMN7." ITU Telecommunications Standardization Sector: H.263 Ad Hoc Group, Feb. 1997.