FGS Enhancement Layer Truncation with Minimized Intra-Frame Quality Variation

Jian Zhou, Huai-Rong Shao, Chia Shen and Ming-Ting Sun

Abstract

This paper proposes an enhancement layer truncation scheme for the Fine-Granularity-Scalability (FGS) video. Our target is to minimize the quality variation of different parts within each frame when the last transmitted enhancement layer is truncated according to the available network bandwidth. This paper proposes an enhancement layer truncation scheme for the Fine-Granularity-Scalability (FGS) video. Our target is to minimize the quality variation of different parts within each frame when the last transmitted enhancement layer is truncated according to the available network bandwidth.

IEEE International Conference on Multimedia and Expo (ICME) 2003

FGS ENHANCEMENT LAYER TRUNCATION WITH MINIMIZED INTRA-FRAME QUALITY VARIATION

Jian Zhou*, Huairong Shao†, Chia Shen† and Ming-Ting Sun*

*Department of Electrical Engineering, University of Washington, Seattle, WA, 98195
†Mitsubishi Electric Research Laboratory, 201 Broadway, Cambridge, MA, 02139

ABSTRACT

This paper proposes an enhancement layer truncation scheme for the Fine-Granularity-Scalability (FGS) video. Our target is to minimize the quality variation of different parts within each frame when the last transmitted enhancement layer is truncated according to the available network bandwidth. We propose to redistribute the bits in the enhancement layer that can only be partially kept with the available bit-budget so that it is able to cover the whole frame area to raise the quality of different parts uniformly. Simulation results confirm the effectiveness of the proposed method in improving the decoded visual quality and reducing the intra-frame quality variation.

Keywords: Fine Granularity Scalability, constant quality, enhancement layer truncation.

1. INTRODUCTION

For today's Internet video streaming applications, one important concern is to intelligently deliver compressed video streams to the end-users with heterogeneous environments. Fine Granularity Scalability (FGS) [1] has been adopted as an amendment to the MPEG-4 standard to address this concern.

The FGS encoder generates two bit streams: one is the base-layer stream to provide the basic visual quality, and the other is the enhancement-layer stream to improve the base-layer quality. The enhancement-layer stream is encoded into several layers with bit-plane coding [2] scheme, and this enables FGS to provide continuous rate-control, in that the enhancement layers can be truncated at any point to achieve the target bit-rate. The corresponded quality of the reconstructed frames is proportional to the amount of enhancement-layer bits received. However, the standard does not specify how to truncate the enhancement layers; it only specifies how to decode the truncated bit stream.

Some research efforts have been introduced on how to truncate the FGS enhancement-layer bit-stream. One simple method, called "even truncation", evenly allocates the available bit-budget to each enhancement-layer frame [3]. This scheme has two defects. (1) The decoded visual quality varies from frame to frame, since the complexity of every frame may be different. We call it inter-frame quality variation. (2) The quality of different parts in a same frame also varies from place to place, since the truncated enhancement layer only covers partial frame area. We call this intra-frame quality variation.

In order to solve the first problem, Zhao [4] uses the nearest feather line to evaluate the importance of each frame, and allocate bits to the enhancement-layer by this criterion. Zhang [5] and Zhao [6] propose to use the optimal rate allocation to truncate the enhancement-layer bit-stream, where the rate-distortion (R-D) curves for each enhancement-layer frame are interpolated during the encoding time to determine the amount of bits that should be truncated. This algorithm minimizes inter-frame quality variation. However, none of the schemes mentioned above has considered the intra-frame quality variation.

For the second defect, Cheong [7] uses water-ring scan order together with selective enhancement to transmit the bit-planes in the area of interest. However, the decoder needs to be modified to decode the water-ring scanned enhancement layers. Another problem is that, for many video sequences with natural scenes, it is hard to define the area of interest, or there might be more than one area of interest. Lim [8] proposes to re-order the enhancement-layer macroblocks according to the quantization values and the coded DCT coefficients of corresponding base-layer macroblock. However, this method does not solve the non-uniform quality in the frame when the enhancement layer is truncated, and the decoder also needs to be modified to decode the enhancement layer.

In this paper, we propose a standard-compatible scheme to reduce the intra-frame quality variation, where the last enhancement layer that can be transmitted for each frame is encoded using the available bit-budget so that it can cover the whole frame area to raise the quality of different parts uniformly. Simulation result confirms the effectiveness of the proposed method.

The rest of the paper is organized as follows. Section 2 presents our enhancement-layer truncation approach.

Simulation results are shown in Section 3, and the conclusion is drawn in Section 4.

2. PROPOSED SCHEME

2.1 Problem Statement

The current MPEG-4 FGS uses a normal scan order to encode the enhancement-layer macroblocks from the upper-left corner down to the bottom-right corner in a frame. As a result, the last bit-plane (layer) of will usually cover part of the frame after the truncation. This is shown in Figure 1. At the decoder side, the upper part covered by the transmitted bit-plane will be enhanced, and the lower part of the frame will not get the enhanced quality, thus the intra-frame quality variation arises.

2.2 Enhancement Layer Truncation

Obviously, if the last bit-plane to be transmitted can cover the whole frame, the quality of the whole frame can be enhanced uniformly. However, the channel bandwidth is often not wide enough to transmit the whole bit-plane.
We solve the above problem by re-encoding the last bit-plane for each frame. Compared with the original last bit-plane, each transcoded block has fewer bits than the original one, but the total amount of bits of the last bit-plane will be the same as the original one. The effect is that, the transmitted bit stream is now able to cover the whole frame area, thus the quality of every block will be uniformly enhanced. The whole process to transcode the enhancement layer of one frame is explained below:
1. Encode the current bit-plane as described in the standard, record the amount of bits generated by each block as R_i where $i = 0,1,...N-1$, (N is the number of block in the frame) and the total amount of bits R_{BP} for the whole bit-plane.
2. If bandwidth allowed, transmit the current bit-plane and go to step 1 to encode the next bit-plane.
3. If not, it means the remaining bit-budget R_{Budget} will not be enough for the whole bit-plane, then the following steps are taken to reduce the number of bits generated in the bit-plane:
   - Shrink the bit budget for each block as:

$$R_i^{'} = R_i - \frac{R_i}{\sum_{i=1}^{N} R_i} \times (R_{BP} - R_{Budget}) \qquad (1)$$

   where $R_i^{'}$ is the new bit-budget to encoder each block. Equation 1 indicates the over-shot bit budget $(R_{BP} - R_{Budget})$ is allocated to each block by its original bits contribution to the whole frame.

- Re-encode the symbols in each block until the new bit-budget $R_i^{'}$ is met. Each enhancement layer block have 64 bits, either "0" or "1", corresponding to the residual errors of DC coefficient to the highest AC coefficients. The encoding procedure with new bit budget means some of the "1" applied to enhance the high frequency DCT coefficients will be dropped.
- Process the next block until the end of the frame.

Our proposed scheme can be combined with other algorithms aiming at reducing the inter-frame quality variation, such as [5], so that both the inter-frame and the intra-frame quality variation can be reduced. This scheme is fully standard compatible, and it only introduces a small portion of extra processing on the enhancement layer, and it can be realized in real-time.

2.3 Rate-Distortion Optimization

In the proposed enhancement layer truncation scheme above, what we do is to drop the "1" bits in the enhancement layer block from which corresponds to the highest AC frequency in the DCT domain. Yet this scheme is not optimized from the rate-distortion point of view. For example, assume the first two consecutive coefficients, 8 and 15, will be encoded in the enhancement layer block, and they can be represented as "1000" and "1111" in the binary form. The MSB, or the first enhancement layer contains two "1". Suppose part of the MSB can be transmitted. If we decide to transmit the bit "1" corresponds to "8", the overall distortion will be 225 in terms of sum of square difference (SSD), while if to transmit the bit "1" corresponds to "15", the overall distortion will be as large as 113 in terms of SSD. On the other hand, to erase the bit "1" related to "15" will generate fewer bits to encode the MSB compared with only erasing the "1" related to "8". So, some balance should be made to decide which "1" bits in the current block should be erased, or, not to be transmitted. The enhancement layer transcoding problem can be generalized as to select some "1" bits from the original block so that the encoded bit-stream will conform to the restricted bit-budget and offer an optimized quality.
Joint rate-distortion optimization method can be utilized to solve this problem. If every block is optimized, then the whole frame is optimized [9]. In one block, for a certain 1, we can minimize the cost function of $J(1) = D(R_i) + 1 R_i$, where R_i is the number of bits used to encode the current block, and $D(R_i)$ is the distortion corresponded to the rate of R_i. As we have mentioned above, when we compute the distortion occurred by erasing the bit "1" in the current bit-plane, the bits associated with the same DCT coefficient in the higher enhancement layer should also be taken into consideration.

In one enhancement layer block, there're 64 bits in one bit-plane. To save the computation, we only consider if the "1" bits will be encoded or erased. The combination of the available erasure pattern will be exponential to the number of "1" in the current block. We can process this using the trellis search method:

- A is the starting of the bit-plane.
- When it reaches the 1st "1" in the bit plane, there're two ways to deal with it, either to keep it as "1" or to modify it as "0", thus two states are generated, namely, "B" and "C". For route "A-B", a cost function can be calculated as $J = \lambda R_1$, where R_1 is the length of the code word to describe the bit string so far. For route "A-C", no cost function is available yet.
- When it reaches the 2nd "1" in the bit plane, 4 routes are generated, namely, "BD", "CD", "BE", "CE". State "E" indicates that this "1" is modified as "0", and state "D" indicates the "1" is kept. For the two routes entering the state "D", one should be discarded, according to the value of $\lambda (R_1 + R_2)$ (corresponding to the route ABD) and $\lambda R_3 + D$ (corresponding to the route ACD), where R_3 is the length of the code word to describe the string of "ACD", and D is the distortion incurred by changing the "1" in position "B" to "0".

One important issue is how to find an approriate λ for the whole frame, so that the generated rate can meet the budget. This can be solved using iteration approaches, such as described in [10]. Another issue is that, according to our observation, the λ value in the consecutive frames are quite similar, thus we can use the λ in the previous frame as the initial λ in the current frame, and this will reduce the computation due to the iterations.

3. SIMULATION RESULT

To validate the effectiveness of our approach, we encoded the Akiyo sequence (CIF format) and compare the performance of "even truncation" with our methods for the enhancement layer. The base-layer is encoded with the quantization parameter of $Q = 31$ for both I frames and P frames. There is no B frame in the sequence. The total available bandwidth for the enhancement layer is 576 kb/s.

Figure 2 shows the PSNR improvement for each frame obtained from our bit-plane truncation algorithm compared with the even truncation algorithm. For the whole sequence, our algorithm which simply drop the "1" bits, as described in Section 2.2, can obtain an average of 0.17 dB improvement. After the R-D optimization, an average improvement of 0.48 dB can be achieved. We use the variance of each macroblock's mean square error (MSE) for the luminance component to measure the intra-frame quality variation. Figure 3 illustrates the intra-frame quality variation for each frame (left) and the deduction of the quality variance (right). Since our method can improve the quality of each block uniformly, for the Akiyo sequence, it reduces the intra-frame quality variation by 26% if we simply drop some "1" bits, and 38% after R-D optimization. Figure 4 shows the decoded frame 61 by even and R-D optimized truncation algorithms. It can be seen that our algorithm has better subjective visual quality.

4. SUMMARY

In this paper, we studied the rate adaptation problem for the FGS enhancement-layers. Specially, we proposed a standard-compatible method that can redistribute the available bit-budget for the last transmitted bit-plane to each block, so that the whole frame can be more uniformly enhanced and the intra-frame quality variation can be reduced. Simulation result proves the effectiveness of the proposed algorithm.

5. REFERENCES

[1] Coding of Audio-Visual Objects – part2 Visual – Amendment 2: Streaming Video Profiles, ISO/IEC 214496-2: 2001 / Amd 2: 2002

[2] W. Li, "Bit-Plane Coding of DCT Coefficients for Fine Granularity Scalability," ISO/IEC JTC1/SC29/WG11, MPEG98/M3989, Oct. 1998.

[3] M. van der Schaar and H.Radha, "A Hybrid Temporal-SNR Fine Granular Scalability for Internet Video", IEEE Trans. on Circuits and System for Video Technology, Vol. 11, No. 3, March 2001, pp. 318-331

[4] L. Zhao, Q. Wang, S. Yang and Y. Zhong, "A Content-based Selective Enhancement Layer Dropping Algorithm for FGS Streaming Using Nearest Feather Line Method", Visual Communications and Image Processing 2002, Proceedings of SPIE, Vol. 4671, pp. 242-249

[5] X.Zhang, A. Vetro, Y. Shi and H. Sun, "Constant Quality Constrained Rate Allocation for FGS Video Coded Bitstreams", Visual Communications and Image Processing 2002, Proceedings of SPIE, Vol. 4671, pp. 817-827

[6] L. Zhao, J. Kim and C. Kuo, "MPEG-4 FGS Video Streaming with Constant-Quality Rate Control and Differentiated Forwarding", Visual Communications and Image Processing 2002, Proceedings of SPIE, Vol. 4671

[7] W. Cheong, K. Kim, G. Park, Y. Lim, Y. Lee and J. Kim, "FGS coding scheme with arbitrary water ring scan order", ISO/IEC JTC1/SC29/WG11, MPEG 2001/m7442, Sydney, July, 2001.

[8]   C. Lim and T. Tan, "Macroblock reordering for FGS", ISO/IEC JTC1/SC29/WG11, MPEG 2000/m5759, March, 2000.

[9]   A. Ortega and K. Ramchandran, "Rate-Distortion Methods for Image and Video Compression", IEEE Signal Processing Magazine, Nov. 1998, pp 23-50

[10]   Y. Shoham and A. Gersho, "Efficient Bit Allocation for an Arbitrary Set of Quantizers", IEEE Trans. on ASSP, vol.36, No.9, Sept. 1998, pp1445-1453
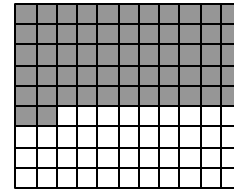
■ The Macroblock that can be transmitted

□ The Macroblock that will be truncated

Figure 1– Effect of FGS enhancement layer bit-plane truncation with normal scan order
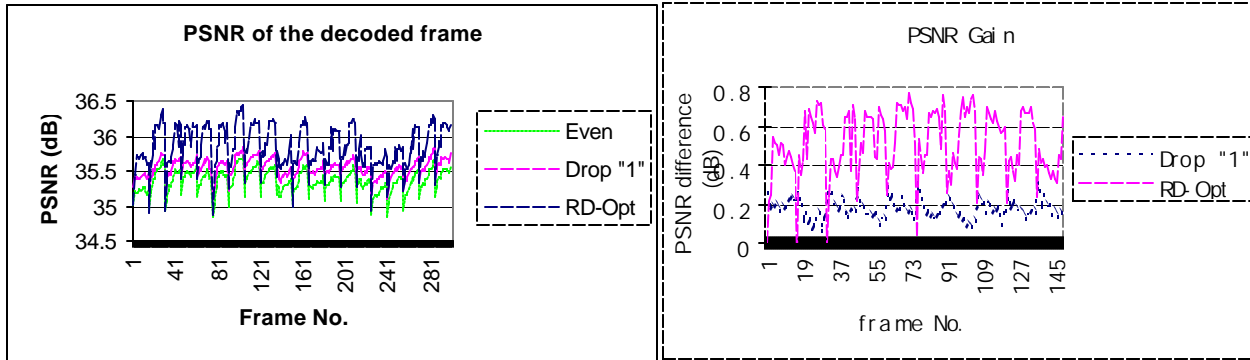
Figure 2– PSNR for each frame (left) and PSNR improvement (right) for each frame in the Akiyo sequence
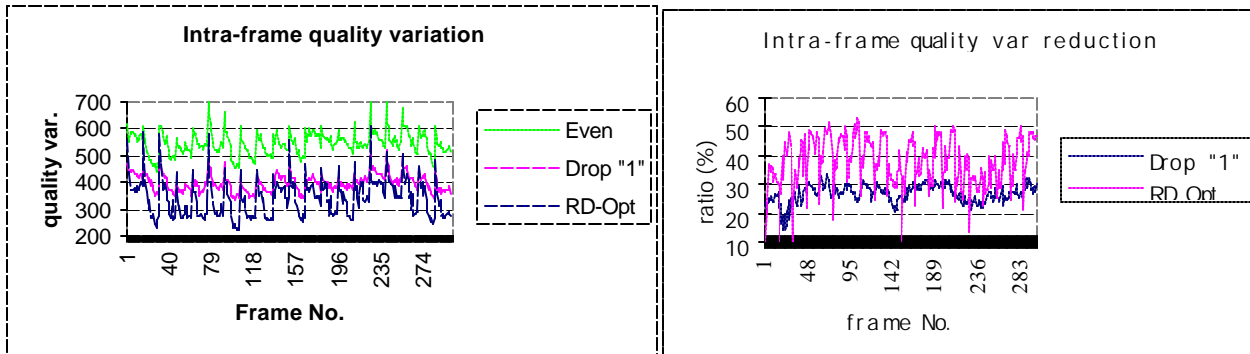

Figure 3– Intra-frame quality variance (left) and variation deduction (right) for each frame in the Akiyo sequence
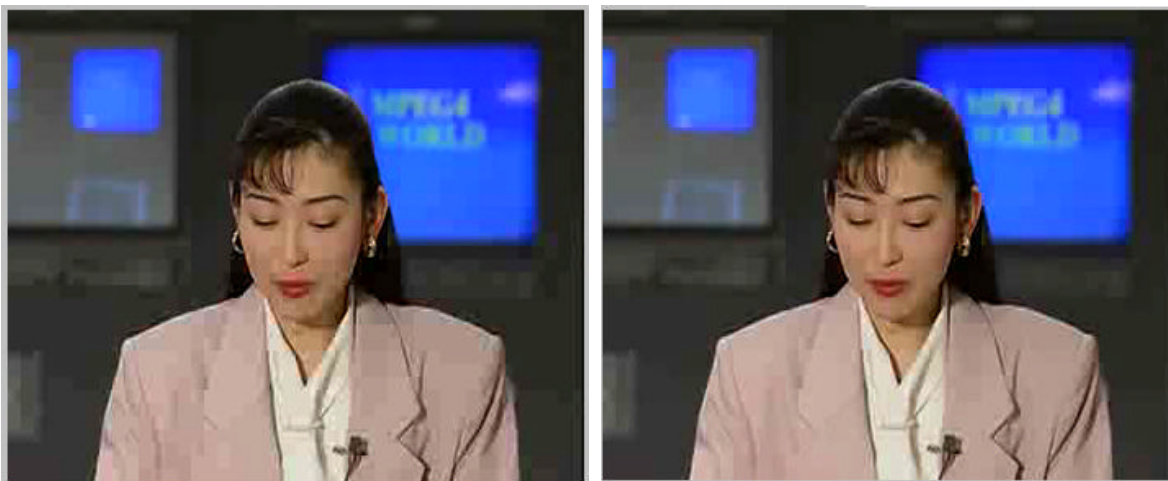

Figure 4– Subjective quality of the decoded frame 61 from even truncation (left) and R-D optimization (right)