OPTIMAL BIT ALLOCATION IN SCALABLE MULTIPLE DESCRIPTION VIDEO CODING FOR PACKET LOSS RESILIENCE

Emrah Akyol[1], A. Murat Tekalp[1,2], and M. Reha Civanlar[1]

[1] College of Engineering, Koc University, Istanbul, Turkey
[2] Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY 14627

ABSTRACT

Scalable multiple description video coders (SMDC) allow for better adaptation of video transmission to varying network conditions with compression efficiency comparable to that of non-scalable MDC coders. Recently, we proposed a new SMDC video coding framework, which offers flexibility in the number of descriptions, redundancy level in each description, and bitrate of each description by post encoding processing only [2]. In this work, we propose a low complexity algorithm to find the best low and high rates for this new framework to strike the optimal trade off between compression efficiency and error resilience at a given packet loss rate. We provide experimental results, by means of network simulations, to compare our method with SMDC using a fixed number of descriptions and fixed redundancy level.

1. INTRODUCTION

Multiple description coding (MDC) can provide robust video communication over unreliable networks such as the Internet or wireless networks by utilizing path/server diversity. MDC addresses the problem of encoding source information using more than one independently decodable and complementary bitstreams, which, when combined, can provide the highest level of quality and when used independently, can still provide lower but acceptable level of quality. This is made possible by introducing redundant bits, which will be discarded if all streams are received. An excellent review of MDC in video communications can be found in [1]

In a recent work [2], we proposed a new wavelet-based flexible SMDC framework with high compression efficiency. The main idea has been to distribute encoded **spatio-temporal code-blocks,** which are obtained after motion compensated temporal filtering and spatial wavelet filtering, to multiple descriptions. Since every code block can be decoded at any given rate, not depending on any other code-block, every code-block is decoded at high bitrate for only one description, where the same code-block is decoded at low bitrate for other descriptions. Different code-blocks, some of which are coded at the high rate and others at the low rate are mixed together to form a description. So if all descriptions are received, decoder uses only code-blocks which are coded at the high bitrate from all the descriptions. On the other hand, if only one description is received we still have an acceptable video quality with code-blocks decoded at low-bitrate. High and low rates determine the amount of redundancy in each description. Hence, it is of interest to determine the optimal high and low bitrates.

Optimizing the number of bits spent on any frame and/or subband is an important issue in video coding since this affects the rate distortion performance of the video coder. Optimal bit allocation for any given bitrate is well studied for both predictive and scalable video coding frame works where the well known Lagrangian Optimization [3] is used to find the optimal bitrates for each frame or macroblock in the rate-distortion sense. In JPEG2000 and JPEG2000-based scalable wavelet video coders rate control is performed with the EBCOT algorithm [4]. In EBCOT, every subband is divided into non-overlapping code-blocks and every code-block is coded independently. Contribution of each block to the total distortion is found by EBCOT and bits from every code-code block are truncated according to their contribution to the overall distortion. The number of bits contributed from each codeblock to overall rate is embedded in the code-block header for fast post compression rate-distortion optimization.

There has been significant amount of early work in optimization of coding parameters in packet loss environment. Rose et al. introduced Recursive Optimal Per-pixel Estimate (ROPE) which optimally chooses the coding mode (inter/intra) in predictive coders considering packet losses and drift problem [5]. Reibman extended

ROPE to MD coding for balanced descriptions, allocating optimum redundancy in each macro-block for a given packet loss rate [6]. Heng et al. [7] considered rate-distortion based optimum MD mode for every macroblock improving MD-ROPE for bursty packet loss scenarios. We propose a similar algorithm for scalable MD coders that optimizes the redundancy level for a given packet loss rate.

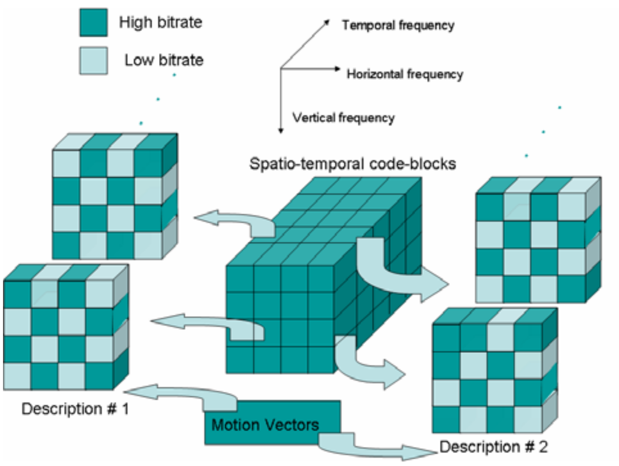

Figure-1: Rate allocation of M=2 descriptions.

In this paper, we find the optimum high and low rates for a given packet loss rate. Since these rates may be determined after encoding, we can adapt redundancy in the SMDC to network conditions on the fly at the server. Hence, the proposed system can respond well to varying network conditions. The paper is organized as follows: we define the rate allocation problem in Section 2. In Section 3, the proposed solution is explained. Section 4 presents comparative results. Conclusions are presented in Section 5.

2. PROBLEM DEFINITION

In SMDC every code-block is decoded with different rates for each description. The assigned extracting bitrates for each code-block are found by performing rate distortion optimization for each block by EBCOT [4] for two bitrates as high bitrate and low bitrate. One code-block decoded at low bitrate belongs to one description and high-rate decoded version belongs to the other description. This process is repeated for all descriptions and code-blocks except the lowest frequency code-blocks in both time and space. Those code-blocks are decoded at high bit-rate in all descriptions since they affect the visual quality more than other code-blocks. If all descriptions are received correctly, all code-blocks are decoded at the high bitrate. Otherwise, received video would still have acceptable quality, where some of the codeblocks are decoded at the low rate. High and low bitrate values and bits spend on motion vector and header information determines the amount of redundancy which is defined as the ratio of the amount of redundant bits that are not used and the amount of bits that are used when all descriptions are received. In this work, we optimize redundancy level with respect to packet loss statistics.

Formally, for N=2 balanced descriptions at a given rate R for each description, and packet loss probability *p,* the problem can be stated as follows: Find the decoding rate of each code-block (R_i) for each description which minimizes the expected total distortion (distortion from source coding and packet loss)

$$D_{total} = D_{packet\_loss} + D_{source\_coding} \qquad (1)$$

such that

$$\sum_{j=1,2} \Big(\sum_{over\_all\_codeblocks} R^i{}_j \Big) \leq 2 \times R = R_{total}$$

3. PROPOSED METHOD

Proposed method is based on Lagrangian Optimization using the information embedded in every code-block. For the case N=2 descriptions, the estimated distortion for code-block *i* can be written as:

$$D_{est}^i = (1-p)^2 D_1^i + (1-p)pD_2^i + p(1-p)D_3^i + p^2 D_4^i \qquad (2)$$

where p is the packet loss probability and D_1, D_2, D_3, and D_4 are distortions respectively when

i) both versions (high rate-low rate decoded) of coded code-block *i* arrive
ii) only low rate code-block arrives
iii) only high rate code-block arrives
iv) none of the code-blocks arrive

The proposed decoder uses the high rate coded code-block, if it exists, otherwise uses the row rate coded code-block. If none of the code-blocks is available, no concealment is performed.

Distortion estimate (2) for a code-block to minimize can be written as

$$D_{est}{}^i = (1-p) \times D_1^i + p \times (1-p) \times D_2^i + p^2 \times D_4^i \qquad (3)$$

In our calculation we ignore the drift problem which is already mitigated by the open loop MCTF structure. We also ignore that the lowest spatio-temporal frequency codeblocks are extracted at high rate in both descriptions. With the assumption of orthogonality of the wavelet filters and motion compensation, total distortion can be written as weighted sum of the code-block distortions as weights are L_2 norms of the wavelet filter coefficients [4], since the distortion in spatial domain will be identical to distortion in the wavelet domain:

$$D_{total} = \sum_{over\_all\_codeblocks} w_i \times D_{est}{}^{i} \quad (4)$$

In JPEG-2000, the optimum rate (in the RD sense) information of that code-block for a given total distortion slope is attached to the code-block in the embedded bitstream for fast post compression rate distortion optimization. To use this information, the problem of interest can be written as : minimize

$$J = D_{total} + \lambda \times R_{total} \quad (5)$$

Since bits spent on motion vectors, other overheads and the distortion where no code-block is available can not be minimized, the problem reduces to investigation of low and high extracting rates for i[th] code-block (i.e., $R_1{}^i$, $R_2{}^i$) which minimizes

$$J = \sum_{1}^{N} [w_i \times ((1-p)\times D_1^i + p\times(1-p)\times D_2^i) + \lambda\times(R_1^i + R_2^i)] \quad (6)$$

where $\sum_{j=1,2} \sum_{i=1}^{N} R_j{}^i = R_{total}$ and N is the number of all code-blocks

Since bit allocation for both high and low bitrate extracted codeblocks are performed with EBCOT, we can safely assume that the total rates for one codeblock is constant, i.e.,

$$R^1{}_i + R_i^2 = R_{i\_total} \quad (7).$$

where R_{i_total} is the rate of the codeblock when all bits are spent on high rate description, found by EBCOT.

Hence, minimizing the total cost for one codeblock would be equivalent to minimizing it over all codeblocks.

$$J = w_i \times ((1-p)\times D_1^i + p\times(1-p)\times D_2^i) + \lambda\times(R_1^i + R_2^i) \quad (8)$$

From the minimization of the expression in (8), we get

$$\lambda_{high} = p \times \lambda_{low} \quad (9)$$

where λ_{high} and λ_{low} respectively correspond to rate-distortion slopes of low and high bitrate coded code-blocks

Note that the rates R^i for each code-block which minimizes RD cost are known for several different slope (λ) values from the information embedded in the bitstream, thanks to the JPEG2000 embedded bitstream structure.

Hence, optimum rates for high and low rate code-blocks can be found by jointly iterating high/ low rates ($R_1{}^i$, $R_2{}^i$) and slopes (λ_{low} , λ_{high}) to satisfy both Eq.7 and Eq.8 using the embedded rate and slope information in the codeblock.

4. RESULTS

We simulate a time-varying channel with NS-2 [8] simulations to demonstrate the performance of our algorithm. The luminescence component of the Foreman sequence in QCIF format is coded with wavelet coder with 3 spatial and 3 temporal decomposition levels for 296 frames at 30fps. Other than the lowest frequency frame in the temporal decomposition, every frame is packetized into one packet with maximum size of 1000 byte. Ever spatial resolution in the lowest frequency frame is put into one packet. All motion vectors for a GOP length 8 are put into a total of 2 packets. Traffic trace files generated in the coder are used in the ns-2 simulation to specify the timing and size of each packet.

There are two senders who have the encoded video and description generator to generate description with any redundancy level. The last hop link is bottleneck link with 100kbps bandwidth and high error rate. Both senders send multiple descriptions over disjoint paths links. Every path from senders to receiver shares one link with 200kbps bandwidth with external traffic. External cross traffic is randomly specified as %50 of the link capacity with exponentially distributed packet sizes and sending intervals. We assume the play-out time is long enough so that all arrived packets are usable.

The simulation time is the two full play time of the video, *T=20sec*. The time period between the time at link changes the packet loss rate and sender sides become aware of that event, the recognition time, is assumed to be t_{rec}=*1sec.*. The rate of each description is set to the bottleneck bandwidth *R=100kbps.*

The proposed system starts with medium redundancy. After a recognition time t_{rec}=*1sec*, it adapts the redundancy level with respect to the packet loss rate %5. At time t=10.sec, after a t_{rec}=*1sec* time from packet loss change , it changes the redundancy level according to the loss rate %20.

For comparison purposes, the performance of a test system with fixed level of redundancy is also simulated.

Since the packet loss rate alternates between %5 and %20 during the simulation, the level of redundancy is fixed according to %15 loss rate.

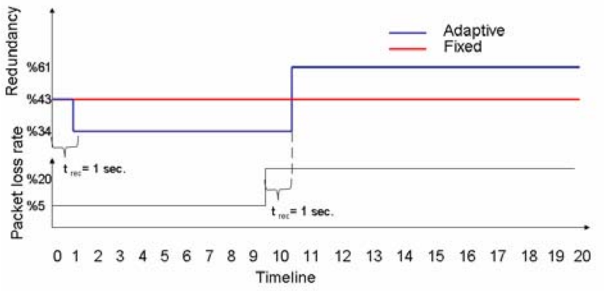

Figure 2: Packet loss and redundancy level in time

Table-1: Low and high bitrates found by our algorithm.

	Packet Loss Rate=%5		Packet Loss Rate=%20	
	Low BR	High BR	Low BR	High BR
Adaptive	50 kbps	148 kbps	75 kbps	124 kbps
Fixed	60kbps	139 kbps	60kbps	139kbps

We compare the proposed system with the test one which has fixed redundancy level through the simulation. Packet loss rates are found by analyzing the ns-2 output trace files of the paths. The results are found by averaging 15 realizations of the simulation.

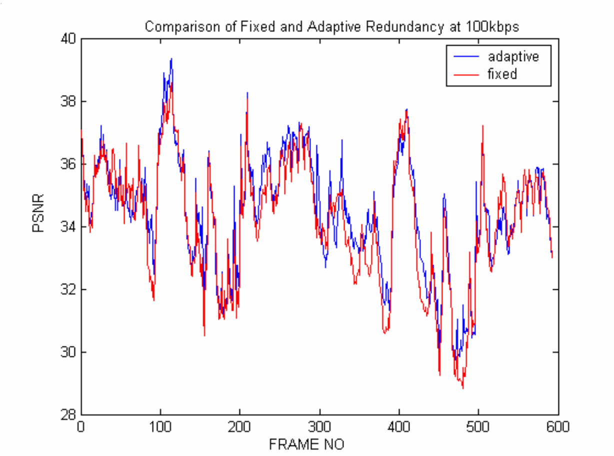

Figure 3: Comparison of adaptive and fixed redundancy

Proposed system with adaptive redundancy level outperforms fixed redundancy by 0.29dB PSNR in the first half and 0.31 dB in the second half of the simulation.

5. CONCLUSION

A rate allocation algorithm is presented for flexible multiple description coding scheme in a packet loss environment. For a given rate and packet loss probability algorithm finds optimum low and high rates for every code-block using the rate-distortion hint tracks found in every code-block header. The low complexity of our algorithm and the flexible nature of SMDC framework enables the varying the amount of redundancy according to network conditions. Comparative results with using fixed redundancy are presented with network simulations

6. REFERENCES

[1] Y. Wang, A. Reibman, and S. Lin, “Multiple description coding for video communications,” *Proceedings of the IEEE*, Jan. 2005.

[2] E. Akyol, A. M. Tekalp, and M.R.Civanlar, “Scalable Multiple Description Video Coding with Flexible Number of Descriptions,” in IEEE Int. Conf. on Image Processing, Genova, Italy, Sept. 2005.

[3] G. Sullivan and T. Wiegand, “Rate-distortion optimization for video compression,” *IEEE Signal Processing Magazine*, vol. 15, Nov 1998.

[4] David S. Taubman and Michael W. Marcellin, “JPEG 2000: Image Compression Fundamentals, Standards, and Practice,” Kluwer International Series in Engineering and Computer Science 2001

[5] R. Zhang, S. Regunathan, and K. Rose, “Video coding with optimal inter/intra-mode switching for packet loss resilience,” *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, June 2000.

[6] A. Reibman, “Optimizing multiple description video coders in a packet loss environment,” in *Packet Video Workshop*, April 2002.

[7] B.A. Heng , J. G. Apostolopoulos, and J. S. Lim, “End-to End Rate-Distortion Optimized Mode Selection For Multiple Description Video Coding,” IEEE Int. Conf. on Acous., Speech and Sign. Proc., Philadelphia, 2005

[8] The Network Simulator NS-2
http://www.isi.edu/nsnam/ns