BALANCED MULTIPLE DESCRIPTION VIDEO CODING USING OPTIMAL PARTITIONING OF THE DCT COEFFICIENTS

Kiran R. Matty, Lisimachos P. Kondi

State University of New York at Buffalo
Department of Electrical Engineering
Buffalo, NY 14260 USA
E-mail: {krmatty, lkondi}@eng.buffalo.edu

ABSTRACT

In this paper, we propose a balanced multiple description video coding scheme which is based on the partitioning of the DCT coefficients. Our scheme splits the single-layered stream generated by a standard coder into two correlated and balanced (equal rates and distortions) substreams. The optimization is in the redundancy-rate-distortion sense using Lagrangian relaxation for optimum allocation of redundancy amongst the blocks in the frame. A greedy algorithm to meet the equal distortions criterion and an algorithm to optimally achieve equal bit rates for the two descriptions is proposed. Our simulation results substantiate our claim of achieving balanced descriptions at high PSNR values, for any specified redundancy or any target bit rate for the encoder. Our design complies to the existing standards, i.e. each description can be independently decoded by any standard decoder.

1. INTRODUCTION

In the recent times, transmission of multimedia over "best-effort" networks such as the Internet has become ubiquitous. But, due to heterogeneity of the networks, congestion and plethora of other reasons, packet losses and delays have become pervasive. If the network supports preferential treatment to some packets, then the use of multiresolution or layered source coding is the obvious choice. But if the networks are oblivious of the contents of the packets and fail to discriminate then the dropping of packets is inevitable. Also packet retransmission (supported by transport protocols like TCP) can be unwarranted due to real-time constraints and absence of feedback channels. Another common problem is network congestion, for example, when packets have to be dropped upon arrival at the switches owing to the fullness of the local buffers. This fosters the need for better source coding algorithms to make the received data meaningful.

Multiple Description Coding (MDC) [1] is an error resilient source coding technique which can be extended to any source (audio, video, data). In this technique a single source of data is represented with several subsets of data popularly called as descriptions such that the source can be approximated from any of these subsets. If these subsets carry identical data over each channel and both channels are operational then half of the information received is redundant, on the other hand if very little dependency exists between the data carried by the subsets, loss of any of the subsets would impair the decoding at the receiver. The primary objective of a MDC coder is to achieve a minimum fidelity reconstruction in the event that only a single channel is operational and should both channels be working, higher fidelity reconstruction should be achievable. Sending information over multiple channels with independent failure events maximizes the odds of receiving some information from at least one of the channels. Some of the desirable qualities of the multiple descriptions are that they should be (a) mutually refining, (b) equally important and (c) relatively independent. The balanced approach i.e. encoding the descriptions at equal rates and distortions, is desirable from the standpoint of transmissions over the Internet, wherein the packets aren't awarded preferential treatment.

2. PROPOSED BMDC SYSTEM

Reibman et al. in [2] proposed a MDC video coder based on rate-distortion splitting wherein the output of the standard video coder is split into two correlated streams. Header information, motion vectors and DCT coefficients with magnitudes above a certain threshold are duplicated into the two descriptions whereas the remaining coefficients are alternated between the two streams. Thresholds for the blocks are computed optimally in rate-distortion sense. At the decoder if both descriptions are received, redundant information is discarded, else the received description is decoded. However, random alternation of non-zero DCT coefficients spawns unbalanced descriptions and also increases the redundancy with the increase in the number of non-zero coefficients.

Our objective is to encode a video sequence into two video substreams with equal bit rates and that are capable of achieving identical PSNR's at the receiver, after being transmitted over two independent channels. Our proposed balanced multiple description coder (BMDC) uses the redundancy-rate-distortion (RRD) [3] criterion at the encoder end, to split the output stream of a standard video coder like the H.263 into two correlated substreams. The encoding algorithm uses the Lagrangian relaxation [4][5] to obtain the optimal threshold values for each block. Two descriptions are created by duplicating (a) header information (b) motion vectors and the (c) DC Coefficients of Intra blocks, after which non-zero AC Coefficients whose magnitudes equal or exceed the threshold values are duplicated into the two descriptions and then the remaining coefficients are split in accordance with our partitioning algorithms to meet the balanced condition. The splitting is done in such a way that when a coefficient value is sent to one of the descriptions, a corresponding zero is sent to the other description. Each description is entropy encoded (run-length coding) [6] independently. In this work, the VLC tables and the syntax of the standard H.263 coder are used.

The operation of an BMDC decoder is influenced by two conditions: 1) if both descriptions arrive at the decoder then the coefficients from the two streams are simply merged into a single stream, that can be decoded by a standard decoder. 2) If either of the descriptions arrives at the receiver then the received stream is simply decoded by a standard decoder like the H.263 decoder. It is noteworthy to observe that the BMDC decoder is oblivious of threshold gymnastics. As a result, no overheads of transmitting threshold values are incurred thereby making our design simple. Also our design complies to the H.263 syntax. Although baseline H.263 is used as the basis of our coder, any motion-compensated DCT based coder could be used.

The concept of redundancy serves as the cornerstone of the MDC paradigm. The $redundancy(\rho)$ is defined as $\rho = (R_1 + R_2 - R^*)/R^*$, where, R^*: Total bits expended to encode a block for the single description coding(SDC) scheme employed using the baseline H.263 encoder, R_1, R_2: Total bits expended in encoding the two descriptions corresponding to a block respectively, such that $R_1 \leq R^*$, $R_2 \leq R^*$. In the redundancy-distortion calculations headers, motion vectors and other overheads are taken into reckoning.

3. ALGORITHM

The redundancy induced and the distortions associated with the descriptions are functions of the threshold. The problem underlying the design of the MDC coder is the selection of the threshold values for optimal allocation of redundancy among the blocks to meet the redundancy budget. Our optimization is at the Group of Blocks (GOB) level. In H.263 with QCIF-sized frames, each GOB comprises of one line of 16 x 16 macroblocks (11 macroblocks). Each macroblock consists of 6 blocks (four luminance and two chrominance 8x8 blocks). Thus each GOB comprises of 66 blocks. DCT coefficients are quantized according to the H.263 standard.

In this paper we discuss two problems. The first problem deals with redundancy allocation amongst the blocks in a GOB and the second problem focusses on meeting the balanced condition. The first problem is formulated as follows: Given that the redundancy-rate-distortion (RRD) function of a block is convex, the threshold values for the optimal allocation of redundancy amongst the blocks within a GOB could be selected by solving the constrained optimization problem formally stated as:

$$\min_{x_i \in \mathbf{X}_i} \sum_{i=1}^{n} D_{\text{block}}(x_i) \tag{1}$$

Subject to

$$\sum_{i=1}^{n} \xi_{\text{block}}(x_i) \leq \xi_{\text{budget}} = \lceil \rho * R^*_{GOB} \rceil, \;\; 0 \leq \rho \leq 1 \tag{2}$$

where $D_{\text{block}}(.)$ and $\xi_{\text{block}}(.)$ are the distortion and the redundancy functions, respectively for an arbitrary block. $D_{\text{block}}(.)$ is the MSE between the dequantized transformed coefficients of either of the descriptions and the unquantized transformed coefficients of a block encoded using the SDC scheme. $\mathbf{X}_i$ is a set of all redundancy-distortion pairs computed offline, corresponding to the i^{th} block for all possible threshold choices. x_i is an operating point corresponding to a optimal threshold for the i^{th} block . For a QCIF-sized image, $n = 66$ (number of blocks in the GOB). R^*_{GOB} refers to the bit rate in bits for a GOB, obtained using the SDC scheme. ξ_{budget} is the redundancy budget constraint in bits on a GOB. ξ_{block} is defined as: $\xi_{\text{block}} = R_1 + R_2 - R^*$ and is measured in bits.

To solve this problem, we consider a RRD framework with MSE as our distortion criterion. Redundancy and distortion measures over the blocks within a GOB are additive. As a result, the constrained problem defined by Eq. (1) and Eq. (2) can be converted to a simpler equivalent unconstrained problem as in Eq. (3) by combining redundancy and distortion through Lagrange multiplier (λ). We solve this unconstrained problem for different positive values of the λ which result in tracing out of the convex hull points of the RRD curve . We pursue the point on the convex hull which yields minimum distortion while meeting the redundancy budget as described in [5].

$$J_{\text{GOB}}(\lambda) = \min_{x_i \in \mathbf{X}_i} \left\{ \sum_{i=1}^{n} D_{\text{block}}\big(x_i(\lambda)\big) + \lambda \sum_{i=1}^{n} \xi_{\text{block}}\big(x_i(\lambda)\big) \right\} \tag{3}$$

$$J_{\text{GOB}}(\lambda) = \min_{x_i \in \mathbf{X}_i} \left\{ \sum_{i=1}^{n} \Big\{ D_{\text{block}}\big(x_i(\lambda)\big) + \lambda \xi_{\text{block}}\big(x_i(\lambda)\big) \Big\} \right\} \tag{4}$$

$$J_{\text{GOB}}(\lambda) = \min_{x_i \in \mathbf{X}_i} \sum_{i=1}^{n} J_{\text{block}}(\lambda) \tag{5}$$

where, $J_{\text{block}}(\lambda) = D_{\text{block}}\big(x_i(\lambda)\big) + \lambda \xi_{\text{block}}\big(x_i(\lambda)\big)$.

In Eq. (3), we have relaxed the budget constraints. The unconstrained problem now becomes the minimization of the lagrangian cost function defined by Eq(4). As seen in Eq. (3), it is possible to express $J_{\text{GOB}}(\lambda)$ as a sum of individual Lagrangian subcosts and then perform the minimization independently for each block by holding λ constant for each block. Then, if the redundancy budget is met for a specific λ, it would ensure that the optimization of the individual Lagrangian costs would yield an optimal redundancy allocation across the whole GOB. The optimal value of the lagrangian multiplier (λ^*) is not known a priori but can be determined using the bisection algorithm [7][5]. At λ^*, $\sum_{i=1}^{n} \xi_{\text{block}}(x_i^*) \simeq \xi_{\text{budget}}$, where x_i^* is the optimal operating point on the Convex hull of the RRD function of the i^{th} block for $i = 1, \ldots, n$.

To meet the balanced condition, we formulate as follows: Given a set of optimal thresholds for splitting the DCT coefficients, partition the coefficients of a block into two subsets subject to the constraint that their bit rates and distortions are identical. Brute force approaches towards the solution of this problem can be an overkill. We describe a simple partitioning algorithm which operates in two phases. Let $\mathbf{C} = \{c_0, c_2, \ldots, c_{63}\}$ be a set of coefficients of an arbitrary block to be split, with the DC coefficient of the Intra blocks and the AC coefficients whose magnitudes equal or exceed the threshold value are duplicated and the remaining AC coefficients are split in accordance with our partitioning algorithm. c_i is the magnitude of the i^{th} coefficient of the block. c_0 is the magnitude of the DC coefficient of the block. Let $\mathbf{C}_1$ and $\mathbf{C}_2$ be the two subsets generated after splitting the coefficients of $\mathbf{C}$ i.e. $\mathbf{C} = \bigcup_{i=1}^{2} \mathbf{C}_i$, Cardinalities $|\mathbf{C}| = |\mathbf{C}_1| = |\mathbf{C}_2| = 64$, for a 8 x 8 block. In the first phase, given the Parseval's theorem, we search for the initial partitions such that the difference in energies of the partitions in the transform domain is minimized to achieve equal

PSNR's at the receiver.

$$\Delta_1 = |\sum_{i=1}^{|\mathbf{C}|} a_i^2 - \sum_{i=1}^{|\mathbf{C}|} b_i^2|. \quad (6)$$

where, $\quad a_i \in \mathbf{C}_1, b_i \in \mathbf{C}_2 \qquad \forall i \in \{1, \ldots, |\mathbf{C}|\}$.
The total energy content of a block is computed, then the coefficients which remain after duplication are assigned to either of the subsets till half the total energy content of the block is reached, after which the remainder of the coefficients are pushed to the other subset thereby equalizing the energy contents of the two subsets. If required, the set $\mathbf{C}$ is permuted to generate a new combination of subsets. We iterate till the energy content of the two subsets are almost equal. Thus, initial partitions of coefficients are obtained. The proposed algorithm is greedy in the sense that it retains the subsets with the smallest difference in energies at each iteration and iterates till the least difference in energies is achieved.
In the second phase, the difference in the bit rates between the two partitions is minimized.

$$\Delta_2 = |R_1 - R_2|. \quad (7)$$

where R_1 and R_2 are the total bit rates associated with the subsets $\mathbf{C}_1$ and $\mathbf{C}_2$ respectively, measured in terms of the number of bits allocated by a standard encoder. Examination of the standard VLC tables for DCT coefficients reveals that replacing a non-zero coefficient by a zero disturbs the run-length which subsequently changes the bit rate. We exploit this observation to achieve identical bit rates for the two descriptions. We start our search by selecting the first non-zero coefficient from either of the initial partitions $\mathbf{C}_1$ or $\mathbf{C}_2$ and parse the other subset to check for its presence by comparing the *absolute* value of the coefficient magnitudes. The number of matches is counted. Depending on the number of matches, coefficients are swapped between the two subsets to generate new combinations of subsets corresponding to each match. If no match were to be found for any selected non-zero coefficient , the next non-zero coefficient in the subset is considered. All the subset combinations assume the same energies as in the initial partitions obtained in the first phase. Of all the combinations, the combination of subsets with the least difference in bit rates is retained. The search is stopped, when the difference in bit rates equates to zero otherwise the combination which achieves lower difference is retained and the algorithm proceeds with the next non-zero coefficient of the selected initial partition till least difference in bit rates is achieved. The algorithm in second phase is optimal given the initial partitioning in the first phase, in the sense that it optimally minimizes Δ_2 over all possible subsets that have the same energy per subset as in the initial partitions obtained in the first phase.

4. RESULTS AND CONCLUSIONS

We used the "Foreman" (10 sec) and "Akiyo" (10 sec) sequences for examining the RRD performances of our proposed algorithm. We compare the performance of our algorithm to the multiple description coding using rate-distortion splitting (MD-RDS) scheme proposed in [2]. Distortion is measured as the average PSNR (dB) over the encoded frames. Bit rate achieved is measured in terms of number of bits allocated by a standard encoder. Quantization parameters are adjusted by the TMN8 rate control algorithm to maintain the overall bit rate of the encoder i.e. SDC coder fixed. The descriptions generated by the BMDC encoder are H.263 syntax compliant and can be easily decoded by the H.263 decoder. We have not used any of the H.263 annexes for the sake of simplicity, but they can very well be supported. We consider the "on-off" model of the channel, which is prevalent over todays Internet.

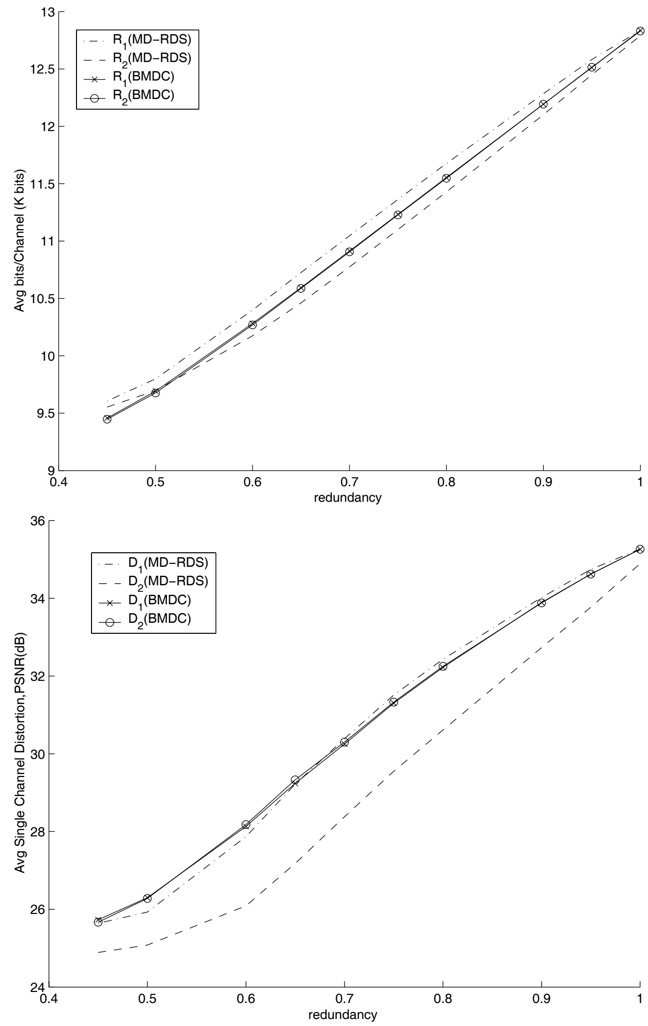


Fig. 1. RRD performance at different ρ values (Foreman), target bit rate for the H.263 encoder (SDC) = 128 Kbps, 2-channel distortion = 35.28 dB.

In the first simulation , we select a target bit rate for the encoder and then modulate the simulation parameter ρ over a range of admissible values. We observe that the bit rates and the PSNR values achieved by the individual descriptions when the MD-RDS encoder is used is disparate over all values of ρ but the those achieved by BMDC are virtually identical over the admissible range of ρ (see Fig. 1). In the second simulation, we present the performance results of our BMDC coder at different target bit rates (see Fig. 2 and Fig. 3) which showcases the robustness of our coder in the face of time-varying channel bandwidths. The reconstruction qualities achieved by MD-RDS are disparate over all target rates but BMDC strives for equal PSNR's at the receiver for all target rates. At any given time, MD-RDS generates one set of partitions per block whereas the BMDC retains only two sets of partitions per block in its pursuit to find the optimal set of partitions, thereby mit-

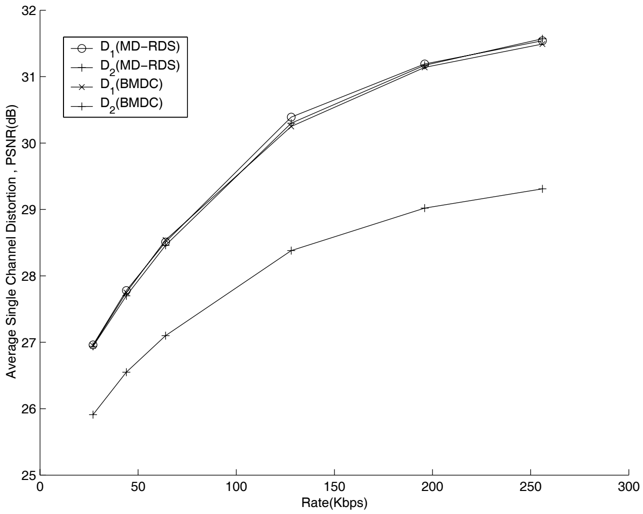


Fig. 2. RRD performance at different target bit rates for the encoder, (Foreman), $\rho = 0.70$.

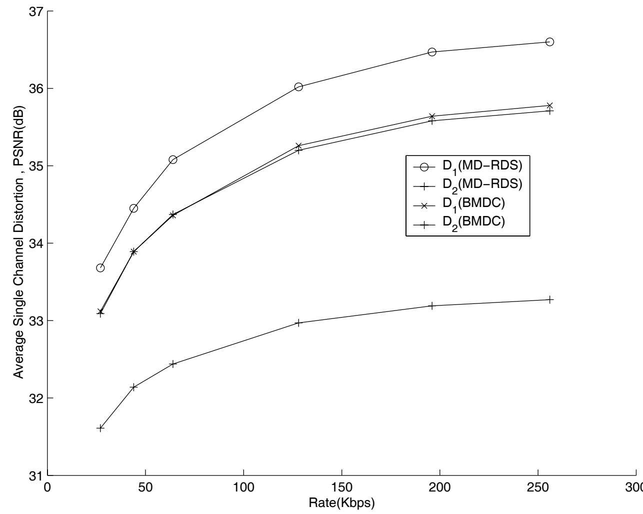
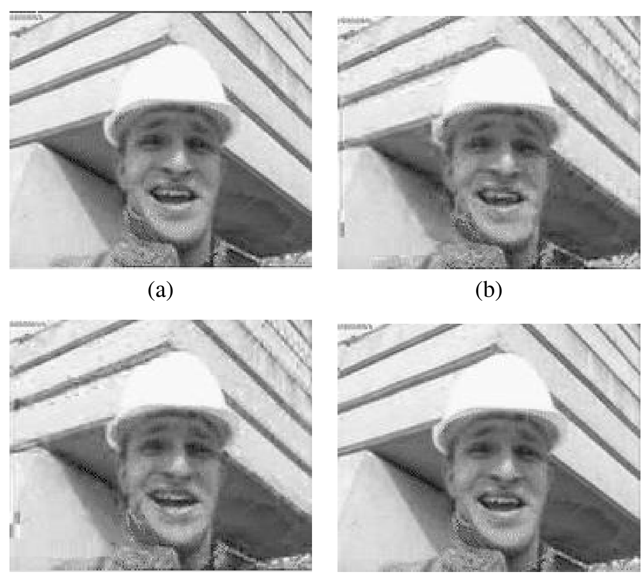


Fig. 3. RRD performance at different target bit rates for the encoder, (Akiyo), $\rho = 0.85$.

igating the spatial costs of employing our partitioning algorithm. Also, we have observed that the time complexity reduces as ρ increases. As more coefficients are duplicated, the number of sets of partitions to be generated to achieve the optimal partitions is reduced. Fig. 4 presents subjective results for the BMDC algorithm, observed for the frame no. 27 of the Foreman sequence for a target bit rate of 128 Kbps for the encoder. Fig. 4(a) is obtained by decoding the single layered bit stream (SDC) generated by baseline H.263 encoder, $\mathbf{R}^*$ represents the corresponding bit rate in bits per pixel (bpp) and D^* is the corresponding distortion. D_0 represents the two channel distortion, i.e. when both descriptions arrive at the receiver.



(a) (b) (c) (d)

Fig. 4. Reconstruction results. Frame no. 27 (Foreman), $\rho = 0.70$ (a) SDC, $\mathbf{R}^* = 0.506$ bpp, $D^* = 35.28$ dB. (b) Reconstructed from description-1, $R_1 = 0.431$ bpp, $D_1 = 30.30$ dB. (c) Reconstructed from description-2, $R_2 = 0.430$ bpp, $D_2 = 30.25$ dB. (d) Reconstructed from both descriptions, $D_0 = 35.28$ dB.

5. REFERENCES

[1] V. K. Goyal, "Multiple description coding: Compression meets the network.," *IEEE Signal Processing Magazine*, vol. 18, pp. 74–93, Sept. 2001.

[2] A.R. Reibman, H. Jafarkhani, Y. Wang, and M. T. Orchard, "Multiple description video using rate-distortion splitting," Tessaloniki, Greece, Oct. 2001, IEEE International Conference on Image Processing (ICIP01).

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

[4] K. Ramchandran and A. Ortega, "Rate-distortion methods for image and video compression," *IEEE Signal Processing Magazine*, vol. 15, pp. 23–50, Nov. 1998.

[5] G. Schuster and A. Katsaggelos, *Rate-Distortion Based Video Compression*, Kluwer Academic Publishers, Boston, 1997.

[6] ITU-T, *H.263 Video Coding for Low bit Rate Communicaitons*, Sept. 1997.

[7] K. Ramchandran and M. Vetterli, "Best wavelet packet bases in rate-distortion sense," in *IEEE Trans. Image Processing*, Apr. 1993, vol. 2, pp. 160–175.