

# UNBALANCED MULTIPLE DESCRIPTION VIDEO CODING WITH RATE-DISTORTION OPTIMIZATION<sup>(\*)</sup>

David Comas, Raghavendra Singh, Antonio Ortega and Ferran Marqués<sup>(\*\*)</sup>

## Abstract

In this paper we tackle the problem of robust streaming of video data over best effort packet networks, such as the Internet. The packet losses and delay, which are commonplace over such networks, can cause severe degradation in the video quality available to the end user. We propose to use multiple description coding (MDC) to protect the transmitted data against packet losses and delay, while also ensuring that the transmitted stream can be decoded with a standard video decoder, such as the H.263 decoder. The video data is encoded into a high resolution, i.e., high quality, video stream (description) using an encoder that produces an H.263 compliant stream. In addition, a low resolution video stream (description) is also generated by duplicating the “important” information from the high resolution video stream. This information includes the headers, the motion vectors and *some* of the discrete cosine transform (DCT) coefficients of the high resolution video stream. The remaining DCT coefficients are set to zero in the low resolution video stream. Hence both video streams are independently decodable by a standard H.263 video decoder. However, only in case of a loss in the high resolution video stream, the corresponding information from the low resolution video stream is decoded, else the received high resolution video stream is decoded. Thus our system is an example of an unbalanced MDC system where the low resolution description is used only in case of losses in the high resolution description. The main contribution of the paper is an optimization algorithm which, given the probability of packet loss, allocates bits to the high resolution and low resolution descriptions, and selects the right number of coefficients to duplicate in the low resolution description, so as to minimize the expected distortion. The MD video coder developed, independently, by Reibman et al. [15] uses a similar rate allocation scheme, with the main difference being that [15] generates balanced descriptions. In this paper we show that the performance of our MD video coder is better than the performance of the MD video coder in [15] for low (<10%) packet loss rate scenarios.

**Keywords:** Multiple Description Coding (MDC), Robust Video Transmission, Rate – Distortion Optimization, Packet Networks.

---

<sup>(\*)</sup> This work was done while the first and second authors were affiliated with USC. The research has been funded in part by the Integrated Media Systems Center, a National Science Foundation Engineering Research Center, Cooperative Agreement No. EEC-9529152, by the National Science Foundation under grant MIP-9804959 and by the CICYT TIC2001-0996 project of the Spanish government

<sup>(\*\*)</sup> D. Comas and F. Marqués are with the Image Processing Group, Department of Signal Processing and Communications, Polytechnic University of Catalonia (UPC), Barcelona (Spain). Email: {comas, ferran}@gps.tsc.upc.es. R. Singh is with IBM India Research Lab. IIT Delhi. New Delhi (India). Email: [raghavsi@in.ibm.com](mailto:raghavsi@in.ibm.com). A. Ortega is with the Signal and Image Processing Institute and Integrated Media Systems Center, Department of Electrical Engineering, University of Southern California, Los Angeles, CA (USA). Email: [ortega@sipi.usc.edu](mailto:ortega@sipi.usc.edu)

## 1. INTRODUCTION

In recent years the volume of multimedia data transmitted over “best-effort” networks such as the Internet has continued to increase while packet losses and delays, due to congestion, routing delay and network heterogeneity, continue to be commonplace. In this paper we address the issue of robust streaming of video data. Video data is usually encoded using predictive encoders, e.g., motion compensation in the standard H.263 [1] and MPEG [2] encoders. These encoders take advantage of the temporal redundancy in the data to achieve high compression performance. However, the main drawback of a predictive coding scheme is that even a single packet loss (or erasure) in the transmitted stream causes decoding errors to propagate through all the samples following the erasure. This severely affects the video quality available at the receiver and motivates the need for robust transmission of video data.

A common approach to limit the length of error propagation in video coders is to restart the prediction loop by periodically inserting Intra coded (non predicted) frames (or macroblocks). A disadvantage of this approach is that there is a loss in coding efficiency due to the frequent restarting of the loop. Moreover the emphasis of this approach is on limiting the error propagation rather than on recovering the lost data. Automatic Repeat Request (ARQ) can be used to retransmit the erased data; however for streaming multimedia applications, especially real time applications, there is a strict time constraint on the transmission of the data, which will limit the number of retransmissions that are possible and thus the overall applicability of ARQ in certain scenarios. Further in a broadcast network scenario that is often used for multimedia data transmission, ARQ can cause NACK implosion at the transmitter [3]. Thus, for streaming video applications local recovery of erasures is often preferable to retransmission. Local recovery at the decoder could be provided with the use of Forward Error Correction (FEC) techniques. In a FEC scheme, redundancy is added to the encoded data so that in case of erasures the redundancy can be used to reconstruct the lost data at the receiver. One drawback of FEC schemes is that they are not as bandwidth efficient as the ARQ schemes; in case of widely changing network conditions, FEC schemes are often designed for the worst case scenario, which usually leads to a waste of precious network resources. Moreover the performance of popular FEC schemes like the FEC channel codes [4] suffers from the cliff effect [5]: for a  $(n, k)$  channel code if the number of errors exceed  $n-k$  then the channel code cannot recover from the channel errors. Hence the performance is constant for up to  $e=n-k$  erasures but then drops very sharply when the actual number of erasures is greater than  $e$ .

An alternative approach for reliable transmission of multimedia data that provides graceful degradation of performance in presence of channel noise is Multiple Description Coding (MDC) [6], [7]. In MDC, two or more independently decodable descriptions of the source are sent to the receiver (Fig. 1). If only description  $S_1$  (or  $S_2$ ) is received, the signal can be reconstructed with acceptable side distortion  $D_{S1}$  (or  $D_{S2}$ ). If both descriptions are received, the distortion obtained at the central decoder,  $D_C$ , is less than or equal to the lowest side distortion; i.e., if  $D_{S1}$  is the lowest side distortion, then  $D_C \leq D_{S1}$ . Thus in an MDC system there are three different decoders, each corresponding to one of the possible loss scenarios. In an ideal MDC channel environment, the channels are independent and data on each channel is either completely lost or received intact. This environment has been studied extensively both

theoretically, [6], [7] and practically, [8], [9], [10], [11], [12], etc. The paper by Goyal [5] provides a good overview of MDC systems.

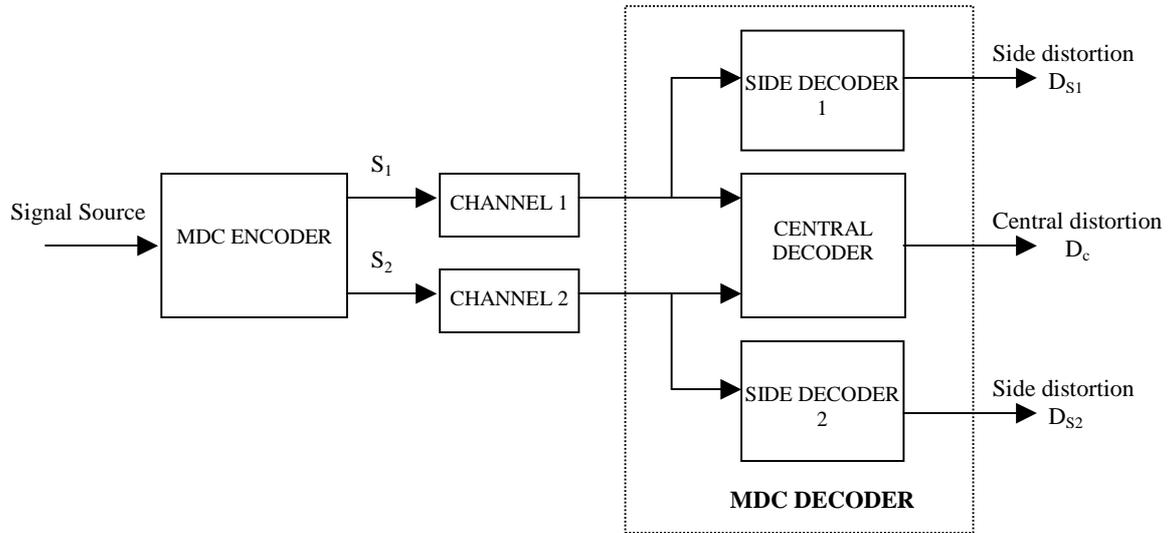


Figure 1 – Generic MDC system. The source is encoded in two descriptions,  $S_1$  and  $S_2$ . They are transmitted through the network over different channels. At the decoder, if only one description is received, the signal can be reconstructed with acceptable side distortion whereas if both descriptions are received the distortion obtained is less than or equal to the lowest side distortion.

In a packet network environment these ideal conditions may not hold true; packet losses can be correlated and only partial data (of either description) may be received at the decoder. Sufficient interleaving of packets of the two descriptions could provide a degree of independence between the packet losses of the descriptions. However there still remains the issue of partially received data, which is especially important for video streaming because of the associated error propagation. There has been limited work, on MD video coders for packet networks. Vaishampayan used MDC scalar quantizers [8] to develop robust image and video coders for packet loss environments. Recently, Reibman has, independently, proposed an MD video coder for packet networks [13], based on a rate allocation principle similar to the one that we propose. One of the novelties of this coder is that in minimizing the expected distortion for a given bitrate, it takes the error propagation into account.

In this paper we propose an unbalanced MDC (UMDC) system for transmission of video data over best effort packet networks. The system is unbalanced because the rate distribution among the various descriptions is not even; hence one description has high rate (high resolution/quality) and the other, low rate (low resolution); i.e., if  $S_1$  is the high resolution (HR) description then  $D_{S1} \leq D_{S2}$  and  $D_C = D_{S1}$ . In the proposed system, the low resolution (LR) description is primarily used as redundancy, to be decoded only when there are losses in the HR description. Most work in MDC has been on balanced systems, i.e., where each description is equally important, but we propose that for the low packet loss rate conditions considered in this paper (below 10%), a UMDC system would be more useful. This is because the overhead in making descriptions balanced, which is particularly significant if the descriptions are to be coded in a standard syntax, would adversely affect the performance of balanced systems for low packet loss rates. Though the UMDC encoder produces unbalanced descriptions, we use smart

packetization, (similar to the one in [14]), to create packets that are equally important; thus from the network viewpoint all packets have equal priority.

In the proposed MDC system, the input video sequence is encoded into a high rate and quality video stream (HR description), using an encoder that produces an H.263 compliant stream. The “important parts” of this HR description are duplicated in a low rate and quality video stream (LR description). The important information includes the headers, motion vectors, and a subset of the DCT coefficients in the HR video stream. The remaining DCT coefficients are set to zero in the LR video stream. At the receiver, if information from the HR description is lost, the corresponding information from the LR description is decoded, else the HR description is decoded. The main advantages of our MD video coder are:

1. Optimal descriptions that minimize the expected distortion for a given probability of packet loss and rate budget are generated.
2. The MD representation is constructed in such a way that both the descriptions are independently decodable by a standard H.263 decoder, i.e., the MD video coder maintains compatibility with the H.263 syntax [1].

The main disadvantage of our work is that currently we are not considering error propagation in our expected distortion formulation. We propose this as part of future work.

The rest of the paper is organized as follows, in Section 2 we discuss other MDC based video transmission systems. In Section 3 we present our encoding algorithm and in Section 4 we compare our system with other MDC systems. We conclude the paper with some thoughts for future work in Section 5.

## 2. RELATED WORK

There has been substantial work in the area of MDC for video transmission over the *ideal* MD channel environment, e.g., [15], [16], [17], [18]. A key challenge in designing MDC techniques that incorporate predictive coding is to avoid the *prediction loop mismatch* problem. When prediction is used, the decoder can only operate properly if it has received the data that the encoder used to generate the predictor, else there is a prediction loop mismatch that leads to poor performance. In an MDC system, if both descriptions were received at the decoder, the best predictor to be used by the central decoder would be one formed from past information produced by the central decoder; i.e., formed by combining information received in both descriptions. However if only a single description was received, the best predictor to use in encoding should be based *only* on data produced by the side decoder corresponding to the description that has been received. Thus ideally, for an MDC based prediction system, there should be three prediction loops, one for each decoder (Fig. 1). Many MD video coders, e.g., [16], [17] and [18] are designed to send redundancy to avoid this mismatch problem.

Reibman et al. [15] proposed an MDC video coder that is similar in principle to our video coder. Descriptions are created by splitting the output of a standard codec; important information (DCT coefficients above a certain threshold, motion vectors and headers) is duplicated in the descriptions while the remaining DCT coefficients are

alternated between the descriptions, thus generating balanced descriptions. The threshold is found in a rate – distortion optimal manner. At the decoder if both descriptions are received then the duplicate information is discarded, else the received description is decoded. This is in principle very similar to our MD video coder, with the main difference being that we duplicate the first  $K$  coefficients of the block and we do not alternate coefficients. The number  $K$  is also found in a rate – distortion optimal framework. The advantage of our method is that its coding efficiency is better than that of [15]. This is because of compliance with the standard syntax of H.263; in our system the efficient end of block (EOB) symbol can be sent after the  $K$ th symbol. Moreover in [15] inefficient runs of zeros are created by alternating DCT coefficients between the descriptions. The disadvantage of our system is that it is unbalanced in nature; hence in case of losses in the HR description there is a sharper drop in performance than in case of losses in either of the balanced descriptions of [15]. However for low packet (<10%) loss scenarios, which are commonplace over the Internet, our system performs better than [13] (a version of [15] extended to packet networks). This is shown in section 4 of this paper.

In [15] the prediction loop mismatch problem that would arise if a description were lost, was not considered. In a later work, Reibman [13], extended this rate distortion optimal splitting method to design an MD video coder for a packet network. For a packet loss environment, prediction loop mismatch could be due to loss of current *and/or* previous data. In [14] the ROPE formulation was developed to calculate the overall distortion of the decoder reconstruction due to quantization, error propagation, and error concealment for a one-layer video coder subject to packet losses. Using this distortion, the best location for intra – blocks was found using rate – distortion optimization. In [13], this optimal mode selection algorithm (i.e., coding a macroblock as inter/intra) was extended to the MD framework to limit the error propagation due to the prediction mismatch. Thus for each macroblock of a frame, given the bitrate and probability of packet loss, the optimal threshold *and* mode was selected; i.e., the redundancy bits were optimally distributed between information needed for local recovery and information needed to limit the error propagation.

In this paper we are proposing a UMD video coder for packet networks. In our system the LR description is used as redundancy to be decoded only in case of losses in the HR description. Thus the central decoder is the same as the HR side decoder, which implies that there is no prediction loop mismatch if there are no losses in the HR description. In case of a loss in the HR description, there is a prediction loop mismatch. This is because at the encoder the information in the HR description is used to predict the next frame, while at the decoder at the point of erasure we will decode the LR description. Thus the prediction for the frame after the erasure will not be from the full HR information but only from the partial information that is available in the LR description. However in the present work we have not considered error propagation due to prediction loop mismatch. The formulation in ROPE, though exact, is computationally intensive and may not be suitable for real time applications. We are currently exploring alternative mode selection methods.

Both our MD video coder and Reibman's ([13] and [15]) are syntax compatible with existing standard codecs. Preserving compatibility with existing standard decoders can affect the performance of an MDC system. For example, it will not be possible to use several techniques, such as MDC scalar quantization or MDC transform coding, while

still preserving compatibility with the standard. However, preserving compatibility may still be useful, because these standard decoders are very commonly used. In particular, if syntax compatibility with standard decoders is preserved, one can think of an MDC system as a wrapper that can use off the shelf encoders and decoders to generate loss – robust transmitted data. Note that by standard compliance we imply that a H.263 compliant decoder can decode either of the descriptions. However in our system, decoding the LR video stream by itself will not give very high quality, as the LR description has been designed only to add robustness to the HR stream. Hence we need a parser, which in case of losses in the HR stream, can extract information from the LR stream and pass it to standard decoder.

In this paper we compare against the MD video coder presented in [15], extended to the packet network environments. We also compare against the video redundancy coding (VRC) mode in H.263+ [19]. In VRC the encoded frames are split into multiple descriptions in a round – robin way and the prediction of a frame in one description is based on the past frames in the same description. In the case of two descriptions, an even frame is predicted from the nearest even frame and an odd frame from the nearest odd frame. Compared to a conventional single description coder there is a significantly lower prediction gain and hence higher redundancy to achieve the same distortion when both descriptions are received. A sync frame is also added periodically to prevent error propagation. The periodicity of the sync frame can be varied to increase robustness but at the expense of additional redundancy. The advantage of VRC is that it is part of a standard and also that it avoids the prediction loop mismatch problem. However in avoiding this problem it adds implicit redundancy that adversely affects its performance.

Note that our UMDC system is not equivalent to the conventional scalable system. In a conventional scalable scheme, e.g., [20], the signal is coded into a low rate base layer and a hierarchy of high rate enhancement layers. An enhancement layer is useful, i.e., it decreases the distortion, if and only if all the layers below it, in the hierarchy, have been decoded. Scalable coding schemes are rate distortion efficient, however due to the dependency of layers their performance decreases sharply in presence of packet losses. On the other hand, in the proposed UMDC scheme the two descriptions are not complementary; the lowest distortion is achieved when the HR description is received. Thus, UMDC as opposed to scalable coding has been designed to minimize the expected distortion at the receiver. Moreover in the proposed UMD system, the important layer is coded at high rate (HR description) while the additional layer is coded at low rate.

### **3. PROPOSED MDC SYSTEM**

The block diagram of the proposed system is shown in Fig. 2. Two descriptions are generated, a high resolution (quality and bitrate) description and a low resolution description. The HR description is obtained by coding the input video sequence by a standard compliant H.263 encoder [21]. The “important parts” of the HR description are duplicated in the LR description. Since the motion vectors and the header information are important, they are transmitted in both descriptions. Moreover for each frame of the video sequence, a select number of high energy discrete cosine transform (DCT) coefficients in a HR block, i.e., a block of the frame in the HR description, are

uplicated in the corresponding LR block. The remaining DCT coefficients in the LR block are set to zero.

For each frame of the video sequence two packets are generated. Each packet contains the headers and the motion vectors, in addition one packet contains the odd Group of Blocks (GOB) of the HR frame, (i.e., the frame in the HR description), and the even GOB of the LR frame. The other packet contains the even GOB of the HR frame and the odd GOB of the LR frame. Thus contents of each packet are independently decodable by a standard H.263 decoder. If both packets are received, then HR GOBs are decoded and LR GOBs are discarded, else the received packet's GOBs are decoded. These equal – importance packets are transmitted over the packet network; virtual independent channels are created by sufficiently interleaving the two packets.

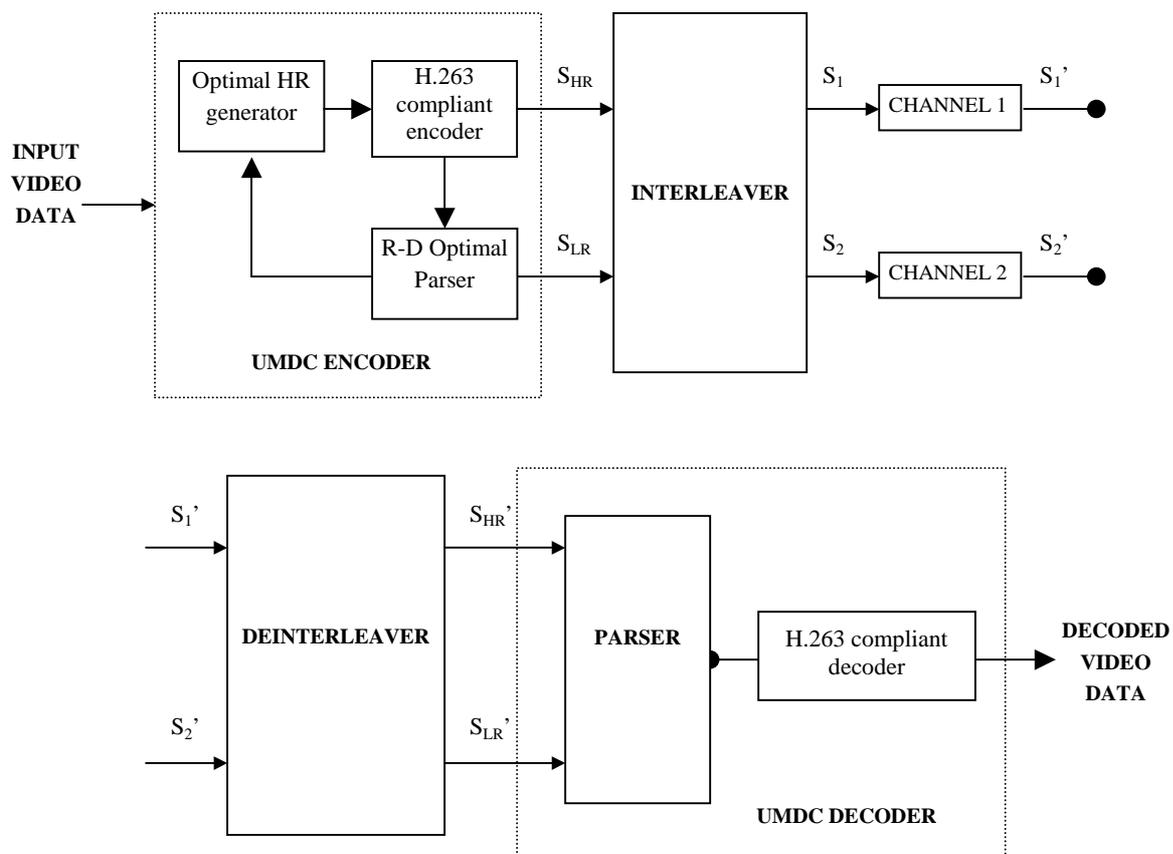


Figure 2 - Proposed UMDC Block Diagram. The input video sequence is coded into two descriptions; one description has high quality and bitrate and is obtained by coding the video sequence by an H.263 compliant encoder. The second description has low quality and bitrate and is obtained by selecting the important information from the high resolution description. This information includes the headers, the motion vectors and a select number of DCT coefficients. There is a feedback to optimally split the total rate into both descriptions according to the probability of packet loss. The descriptions are packetized, 2 packets per frame, such that each packet is of equal size and importance. Virtual independent channels are created over a packet network by sufficiently interleaving the two packets. If both packets are received then the HR information is parsed from the packets and the LR information is discarded, else the contents of the received packets are decoded.

The main contribution of this work is that the descriptions are generated in a rate – distortion optimal framework; i.e., given the probability of packet loss and the total available rate,  $R_{TOT}$ , the MD video coder generates HR and LR descriptions that minimize the expected distortion. This involves finding the rate allocation,  $R_{HR}$ , for the HR description, coding the HR description, and parsing the resulting HR video stream to select the right number of coefficients to duplicate in the LR description. In order to formulate this optimization problem, let us assume that there are  $N$  frames in the video sequence, with  $M$  macroblocks per frame, and let  $d_{HR}^{i,j}$  and  $d_{LR}^{i,j}$  represent the distortion in the  $j$ th macroblock of  $i$ th frame of the HR and LR descriptions, respectively. Let  $E$  represent the set of all macroblocks in the even GOBs of a frame and  $O$  represent the set of all macroblocks in the odd GOBs of a frame. Let  $r_{HR}^i$  and  $r_{LR}^i$  represent the  $i$ th frame rate for the HR and LR descriptions respectively. Given the packetization policy, the expected distortion can be written as,

$$E(D) = \sum_i^N (1 - p_i)^2 \cdot \sum_j^M d_{HR}^{i,j} + \sum_i^N p_i \cdot (1 - p_i) \left( \sum_{j \in E}^M d_{HR}^{i,j} + \sum_{j \in O}^M d_{LR}^{i,j} \right) + \sum_i^N p_i \cdot (1 - p_i) \left( \sum_{j \in O}^M d_{HR}^{i,j} + \sum_{j \in E}^M d_{LR}^{i,j} \right) \quad (1)$$

where  $p_i$  is the probability of packet loss for frame  $i$ . In the above formulation we are ignoring the case when both packets of a frame, are lost. The objective is to minimize the above expected distortion under the constraint,

$$R_{HR} + R_{LR} = R_{TOT} \quad (2)$$

i.e.,

$$\sum_N r_{HR}^i + \sum_N r_{LR}^i = R_{TOT} \quad (3)$$

Solving this constrained optimization problem can be extremely complex. Due to the predictive nature of the video coder,  $d_{HR}^{i,j}$ ,  $r_{HR}^j$  depend on  $d_{HR}^{i-1,j}$ ,  $r_{HR}^{i-1}$  (true also for LR description). Further in our MDC system  $d_{LR}^{i,j}$  is also dependent on  $r_{HR}^i$  because the low resolution description is generated from the high resolution description.

We make the assumption that each frame is coded independently at a bitrate,  $r_{TOT}^i$ , where,

$$\begin{aligned} \sum_N r_{TOT}^i &= R_{TOT} \\ r_{HR}^i + r_{LR}^i &= r_{TOT}^i \end{aligned} \quad (4)$$

Thus, the constrained optimization can be solved independently for each frame. Rewriting the constrained problem as an unconstrained minimization problem using the Lagrangian multiplier  $\lambda$  [22], where the objective is to minimize, for each frame  $i$ , the cost function,

$$J(D) = (1-p_i)^2 \cdot \sum_j^M d_{HR}^{i,j} + p_i \cdot (1-p_i) \left( \sum_{j \in E} d_{HR}^{i,j} + \sum_{j \in O} d_{LR}^{i,j} \right) + p_i \cdot (1-p_i) \left( \sum_{j \in O} d_{HR}^{i,j} + \sum_{j \in E} d_{LR}^{i,j} \right) + \lambda \cdot (r_{HR}^i + r_{LR}^i - r_{TOT}^i) \quad (5)$$

With this formulation, allocation is done independently for each frame based on the budget obtained from TMN8 [23]. Therefore, in what follows we can ignore the frame index  $i$  when stating the objective function.

In this present work we do not take error propagation into account; we are currently exploring ways of incorporating the error propagation in this formulation while keeping the computational cost reasonably low for real – time applications. Thus the distortion,  $d_{HR}^j$ , of the  $j$ th macroblock in the HR description is a function of the quantization parameter  $Q^j$ , while  $d_{LR}^j$  is a function of  $Q^j$  and  $k^j$ , the number of DCT coefficients duplicated, i.e., not set to zero in *all* 8x8 blocks in the macroblock  $j$ . We assume that all the blocks in a macroblock employ the same value of  $k^j$ . This is reasonable since most of the processes done in the encoder treat macroblocks as single entities (e.g., motion estimation, header generation, etc). The optimization can be then written as,

$$\min_{\mathbf{Q}, \mathbf{k}} \left[ (1-p)^2 \cdot \sum_j^M d_{HR}^j(Q^j) + p(1-p) \left( \sum_{j \in E} d_{HR}^j(Q^j) + \sum_{j \in O} d_{LR}^j(Q^j, k^j) \right) + p(1-p) \left( \sum_{j \in O} d_{HR}^j(Q^j) + \sum_{j \in E} d_{LR}^j(Q^j, k^j) \right) + \lambda \cdot (r_{HR}(Q^j) + r_{LR}(Q^j, k^j) - r_{TOT}) \right] \quad (6)$$

where  $\mathbf{Q}$  and  $\mathbf{k}$  are the sets of quantization steps and DCT coefficients duplicated in LR for each block  $j$  in the current frame, respectively. In our previous work, [24], we developed an algorithm for finding the optimal  $\mathbf{k}$ , given  $\mathbf{Q}$  and  $p$ . We assume that only the first  $k^j$  DCT coefficients, along the zig – zag scan, are duplicated. To reduce the search complexity, the admissible values of  $k^j$  are restricted to be 0, 1, 2, 4, 8, 10, 12, 16, 32 or 64.

We use mean square error (MSE) as our distortion metric. Since DCT is an orthogonal transform, the distortion in a block of a frame can also be written as a function of the corresponding transform coefficients. Let  $l$  denote the index of the luminance block in the current macroblock  $j$  ( $l \in (1..4)$ ). Thus  $C_{l,j}^n$  is the  $n$ th coefficient in block  $l$  from the current macroblock  $j$ , and  $\tilde{C}_{l,j}^n$ , its quantized value, dependant on the quantization step used ( $Q^j$ ). We then can define the distortion in the HR and LR macroblocks, as follows:

$$d_{HR}^j(Q^j) = \sum_{l=1}^4 \sum_{n=1}^{64} (C_{l,j}^n - \tilde{C}_{l,j}^n(Q^j))^2 \quad (7)$$

$$d_{LR}^j(Q^j, k^j) = \sum_{l=1}^4 \left( \sum_{n=1}^{k^j} (C_{l,j}^n - \tilde{C}_{l,j}^n(Q^j))^2 + \sum_{n=k^j+1}^{64} (C_{l,j}^n)^2 \right) \quad (8)$$

The rate and distortion of a macroblock for each possible  $k^j$  is calculated directly in the H.263 codec. We ignore headers, motion vectors and other side information since they will be present in both descriptions.

The proposed rate – distortion optimal algorithm for generating the descriptions can be summarized into the following steps:

For each frame of the input sequence

- a. Find the required frame rate  $r_{TOT}$  using the TMN8 rate control algorithm
- b. For the allowed set of  $\mathbf{Q}$  and  $\mathbf{k}$ , find all the possible  $d_{HR}(Q^j)$  and  $d_{LR}(Q^j, k^j)$ .
- c. Minimize the expected frame distortion (1) under the rate budget,  $r_{TOT}^i$ , constraint (6).
- d. Generate HR and LR according to the optimal  $\mathbf{Q}$  and  $\mathbf{k}$  found.

#### 4. SIMULATIONS AND RESULTS

The reported results are expressed in terms of the Signal to Noise Ratio (SNR) of the luminance components of the first 100 frames of the two QCIF (176x144 pixels) test sequences, namely, *Akiyo* and *Coastguard*. We compare the performance of our UMD coder (UMDC) with the MD–RDS coder proposed by Reibman et al. [15]. The work in [15] has been extended to a packet network environment, similar to the work in [13] except that optimal mode selection is not considered in the present work. Both systems are coded at the same total bitrate,  $R_{TOT}$ , and each frame is packetized into two packets. In the UMDC case the packetization is as explained in the previous section, whereas in the MD–RDS case, each packet contains a description. Thus while our system creates packets of equal size and importance by combining portions of the HR and LR descriptions, in MD-RDS each description can be packetized separately due to the balanced nature of the descriptions. In both cases if both descriptions are completely lost we replace the lost information with the spatially corresponding information from the previous frame. If this situation happens in the first frame, where no information is available, the lost macroblocks are set to their statistical mean value. We consider random losses with identical loss sequences being injected in both systems.

Figure 3 and 4 show the results for the sequence *Akiyo* coded at different bitrates. The frame rate is 30fps and only the initial frame of the stream is intra coded. The results show that for lossy conditions, MD coding method easily outperforms the single description (with no redundancy) method *SD*. Among the MD methods, UMDC performs better than MD – RDS for low packet loss conditions; this improvement in performance increases with increasing bitrate. This can be explained by the fact that MD – RDS has an implicit redundancy because it alternates coefficients in order to make the descriptions balanced. Alternate non – zero coefficients adversely affect the entropy coder of standards like H.263. This implicit redundancy increases with the increase in the number of non – zero coefficients.

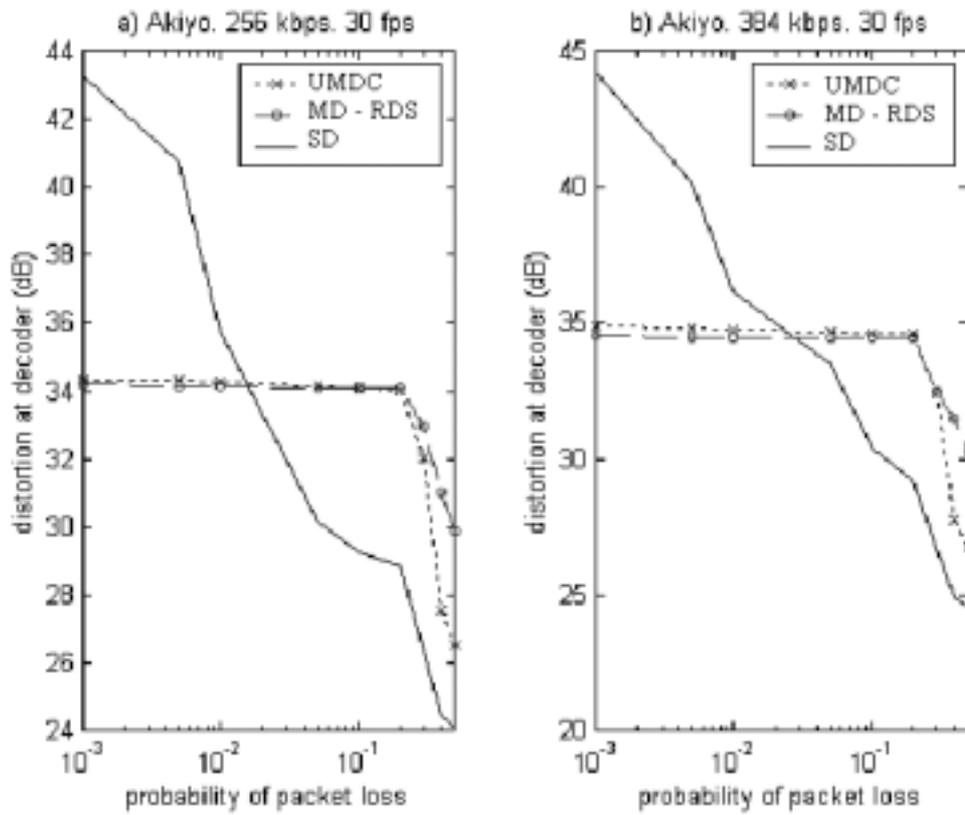


Figure 3 – UMDC vs. MD – RDS and SD. Akiyo. 256(a) and 384(b) kbps.

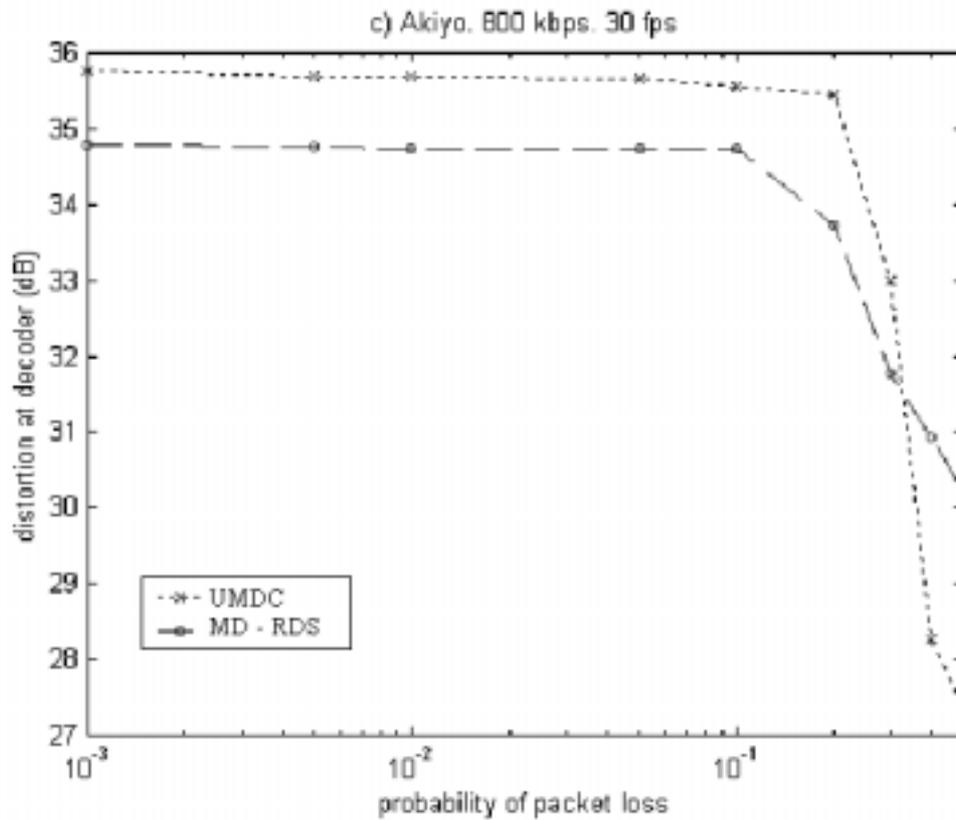


Figure 4 – UMDC vs. MD – RDS. Akiyo. 800 kbps

Plots in Figure 5 shows the results for the sequence *Coastguard*.

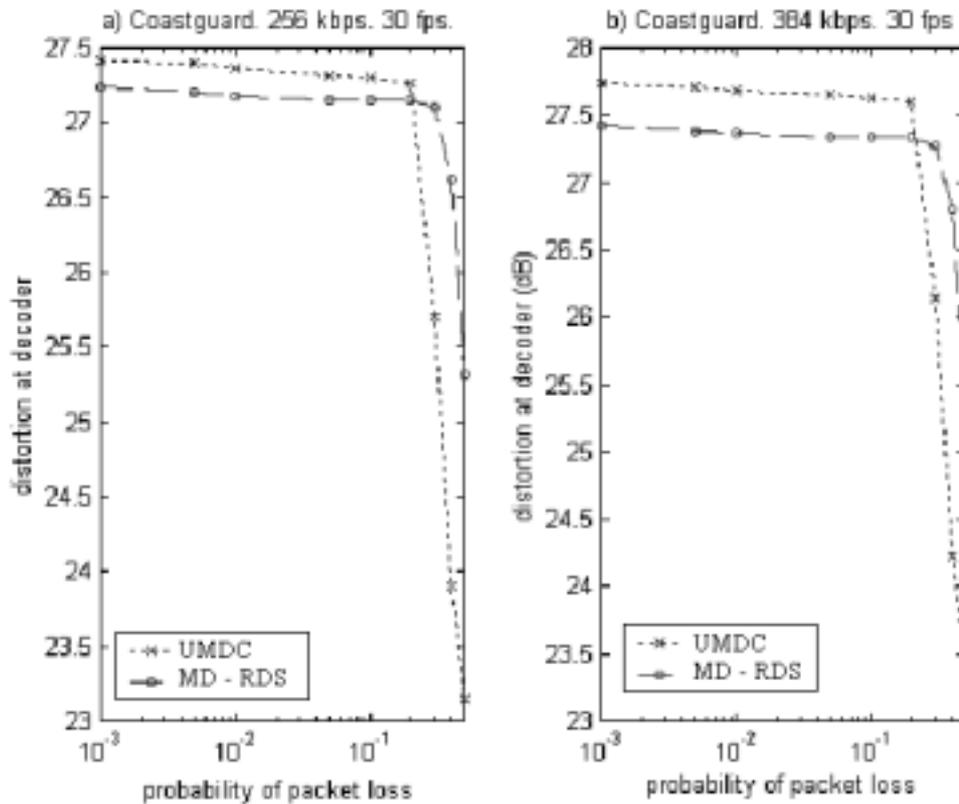


Figure 5 – UMDC vs. MD – RDS. *Coastguard*. 256 (a) and 384(b) kbps.

Figure 6 shows the comparison of both systems when only the first frame is coded in intra frame mode (Figure 6a) and when an intra frame is transmitted every 10 frames (Figure 6b). In this last case, a larger number of DCT coefficients are different from zero, since Intra frames contain a larger number of non – zero transform coefficients than inter frames. Hence the coding efficiency of the UMD coder is much better than that of MD – RDS. The maximum gain over MD – RDS when an intra frame is inserted every 10 frames is around 1.3dB whereas in the case where only the first frame is coded without prediction, the gain is about 0.8 dB. The results also show that MD methods perform better than vanilla Intra update methods. The test sequence used is *Akiyo*, coded at 512 kbps.

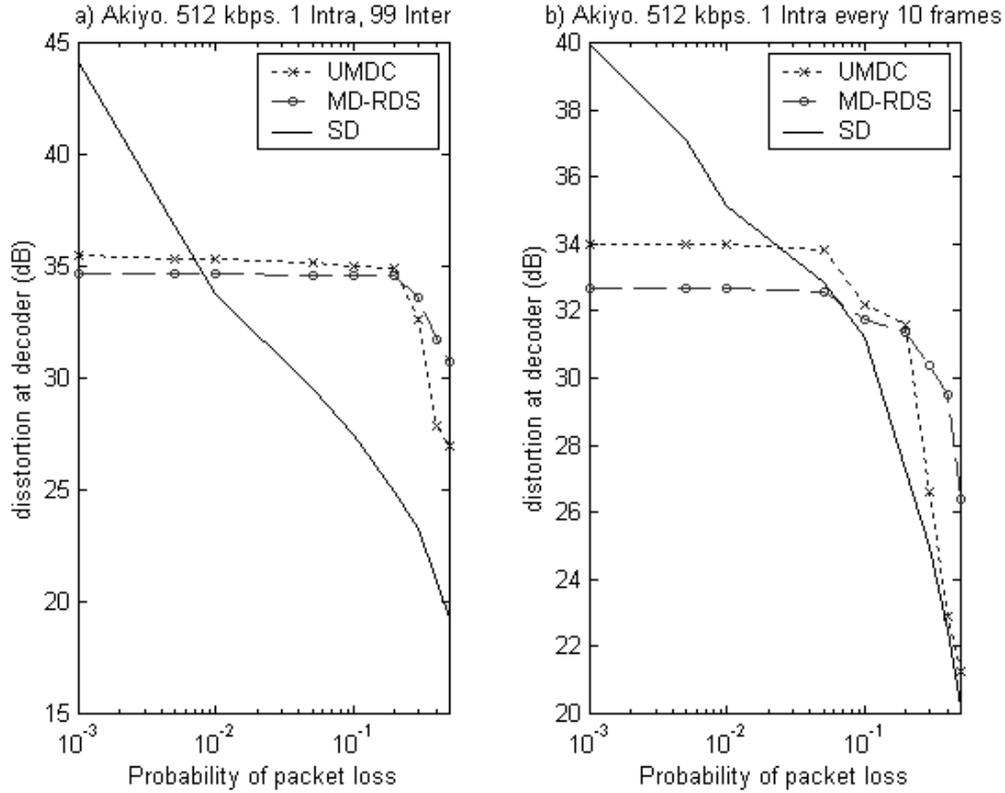


Figure 6 – UMDC vs. MD – RDS and SD. Akiyo. 512 kbps. 30 fps

In the table 1 we compare UMDC with VRC. For VRC we sent two packets per frame, where one packet contained only the even GOBs and the other packet odd GOBs. Akiyo sequence is coded at 30kbps with frame rate 10fps. Again UMDC does very well for low packet loss scenarios.

Probability of packet loss	3%	5%	10%
UMDC	34.05 dB	34.01 dB	33.89 dB
MD – RDS	33.91 dB	33.87 dB	33.8 dB
VRC	33.7 dB	33.4 dB	32.5 dB

Table 1 – UMDC vs. MD – RDS and vs. VRC. Akiyo. 30 kbps. 10 fps.

## 5. CONCLUSIONS AND FUTURE WORK

We have shown that the proposed UMDC system enhances the robustness of video coders to packet losses using a very small amount of extra computational resources and being compliant with the existing video decoder standard syntax. The algorithm takes advantage of the processes done in the encoder and then, by just pruning coefficients in a rate – distortion framework, generates a low resolution stream to be used in case the main stream (HR) is lost.

One of the main emphases in our work was to design an MDC video system being compatible with existing codecs. As previously commented, better results could be obtained by removing this constraint.

Comparing our proposed system to MD – RDS we have seen that the main benefits of UMDC are due to the unbalanced nature of the proposed scheme: by coding runs of zeroed DCT coefficients, the coding efficiency obtained is larger than the one obtained by alternating the transform coefficients. As part of our future work we would like to keep on studying different options of unbalanced multiple description coding schemes and compare them against balanced ones.

In future work we would also like to take error propagation into account, as in [14], but since the generation of the low resolution stream is done frame by frame an appropriate model of the rate distribution among frames prior to any encoding should be found.

Another future work we would like to do is to use the low resolution sequence not just as plain redundancy, but also as a refiner of the high resolution description.

## ACKNOWLEDGEMENTS

The authors would like to thank both the reviewers for their helpful suggestions.

## REFERENCES

- [1] H.263 Video Coding for Low Bit Rate Communication. ITU-T. September 1997.
- [2] Motion Pictures Experts Group (MPEG). ISO/IEC JTC1/SC2/WG11. 1992
- [3] R. Rejaie, M. Handley and D. Estrin. “*Quality Adaptation for Congestion Controlled Video Playback over the Internet*”. Proc. of ACM SIGCOMM’99. Cambridge, MA., September 1999.
- [4] S. Lin and J.D.J Costello. “*Error control coding: Fundamentals and applications*”. Prentice-Hall. 1983.
- [5] V.K. Goyal. “*Multiple Description Coding: Compression Meets the Network*”. IEEE Signal Processing Magazine, vol. 18, n°. 5, page(s) 74 – 93. September 2001
- [6] L. Ozarow. “*On a source-coding Problem with two Channels and three Receivers*”. Bell System Technical Journal, vol. 59, page(s) 1909 – 1921. December 1980.
- [7] A. A. El-Gamal and T. M. Cover. “*Achievable rates for multiple descriptions*”. IEEE Trans. on Information Theory, vol. 28, page(s) 851 – 857. November 1982.
- [8] V.A. Vaishampayan. “*Design of multiple description scalar quantizers*”. IEEE Trans. on Information Theory, vol. 39, page(s) 821 – 834. November 1993.
- [9] V.K. Goyal, J. Kovacevic, R. Aream and M. Vetterli. “*Multiple Description Transform Coding of Images*”. IEEE International Conference on Image Processing (ICIP98), Chicago, IL, October 1998
- [10] Y. Wang and M.T. Orchard and A.R. Reibman. “*Optimal Pairwise Correlating Transforms for Multiple Description Coding*”. IEEE International Conference on Image Processing (ICIP98), Chicago, IL, October 1998
- [11] R. Puri and K. Ramchandran. “*Multiple Description Source Coding using Forward Error Correction Codes*”. Proc. of 33<sup>rd</sup> Asilomar Conference on Signals and Systems. 1999.

- [12] W. Jiang and A. Ortega. “*Multiple Description Coding via Polyphase Transform and Selective Quantization*”. In Proc. of Visual Communic. and Image Proc., VCIP '99, San Jose, Ca, January 1999
- [13] A. Reibman; “*Optimizing multiple description video coders in a packet loss environment*”. Proc. of Packet Video Workshop, Pittsburgh, USA, March 2002.
- [14] R. Zhang; S. L. Regunathan; K. Rose; “*Video coding with optimal inter/intra mode switching for packet loss resilience*”. IEEE Journal on Selected Areas in Communications, vol. 18 – 6, June 2000, Page(s): 966 –976
- [15] A.R. Reibman; H. Jafarkhani.; Y. Wang; M. T. Orchard, “*Multiple description video using rate-distortion splitting*”. IEEE International Conference on Image Processing (ICIP01), Tesseloniki, Greece, October. 2001.
- [16] V.A. Vaishampayan and S. John. “*Interframe Balanced Multiple Description Video Compression*”. Proc. of Packet Video Workshop. 1999.
- [17] A.R. Reibman; H. Jafarkhani; Y. Wang; M. T. Orchard; R. Puri; “*Multiple Description Coding for Video using motion compensated prediction*”. IEEE International Conference on Image Processing (ICIP99), Kobe, Japan, October. 1999.
- [18] Y. Wang and S. Lin. “*Error resilient video coding using multiple description motion compensation*”. Proc. of 2001 IEEE Fourth Workshop on Multimedia Signal Processing (MMSP01), Cannes, France, October 2001
- [19] S. Wenger; “*Vide Redundancy Coding in H.263+*” Proc. of Workshop on Audio Visual Services for Packet Networks, Scotland 1997.
- [20] M.van der Schaar; H. Radha. “*A hybrid temporal – SNR fine – granular scalability for intertet video*”. Circuits and Systems for Video Technology, IEEE Transactions on, Vol 11, Issue:1, March 2001, Page(s): 318-331
- [21] M. Gallant; F. Kossentini; Department of Electrical Engineering. University of British Columbia, Canada. H.263 TMN video encoder. Version 3.0. May 1997.
- [22] A. Ortega; K. Ramchandran; “*Rate-Distortion methods for image and video compression*”. IEEE Signal Processing Magazine Volume: 15 – 6, Nov. 1998, Page(s): 23 – 25
- [23] Ribas – Corbera, J.;Lei, S. “*Rate control in DCT video coding for low-delay communications*”. Circuits and Systems for Video Technology, IEEE Transactions on, Vol:9, Issue:1, Feb.1999, Page(s): 172 –185
- [24] D. Comas; R. Singh; A. Ortega. “*Rate-Distortion Optimization in a Robust Video Transmission based on Unbalanced Multiple Description Coding*”. Proc. of 2001 IEEE Fourth Workshop on Multimedia Signal Processing (MMSP01), Cannes, France, October 2001