EURASIP Journal on Applied Signal Processing 2003:1, 81–90

Unbalanced Multiple-Description Video Coding with Rate-Distortion Optimization

David Comas

Image Processing Group, Department of Signal Theory and Communications, Technical University of Catalonia (UPC), 08034 Barcelona, Spain
Email: comas@gps.tsc.upc.es

Raghavendra Singh

IBM India Research Lab., Indian Institute of Technology (IIT) Delhi, New Delhi, India
Email: raghavsi@in.ibm.com

Antonio Ortega

Signal and Image Processing Institute and Integrated Media Systems Center, Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089-2564, USA
Email: ortega@sipi.usc.edu

Ferran Marqués

Image Processing Group, Department of Signal Theory and Communications, Technical University of Catalonia (UPC), 08034 Barcelona, Spain
Email: ferran@gps.tsc.upc.es

Received 7 May 2002 and in revised form 11 November 2002

We propose to use multiple-description coding (MDC) to protect video information against packet losses and delay, while also ensuring that it can be decoded using a standard decoder. Video data are encoded into a high-resolution stream using a standard compliant encoder. In addition, a low-resolution stream is generated by duplicating the relevant information (motion vectors, headers and some of the DCT coefficient) from the high-resolution stream while the remaining coefficients are set to zero. Both streams are independently decodable by a standard decoder. However, only in case of losses in the high resolution description, the corresponding information from the low resolution stream is decoded, else the received high resolution description is decoded. The main contribution of this paper is an optimization algorithm which, given the loss ratio, allocates bits to both descriptions and selects the right number of coefficients to duplicate in the low-resolution stream so as to minimize the expected distortion at the decoder end.

Keywords and phrases: multiple-description coding, robust video transmission, rate-distortion optimization, packet networks.

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, for example, 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 intracoded (nonpredicted) 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

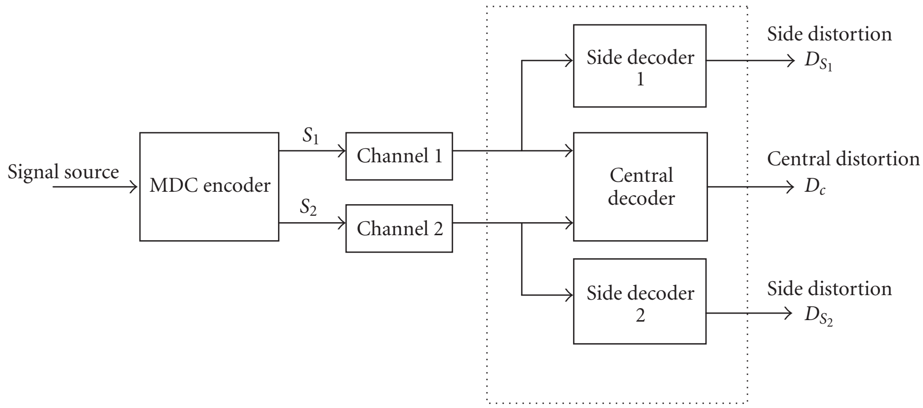


FIGURE 1: Generic MDC system. The source is encoded in two descriptions: S_1 and S_2. They are transmitted through the network over different channels. At the decoder, if only one description is received, the signal can be reconstructed with acceptable side distortion whereas if both descriptions are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

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, consequently, 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 a negative acknowledgement (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 an 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 an (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 (Figure 1). If only description S_1 (or S_2) is received, the signal can be reconstructed with acceptable side distortion D_{S_1} (or D_{S_2}). If both descriptions are received, the distortion obtained at the central decoder D_C is less than or equal to the lowest side distortion; that is, if D_{S_1} is the lowest side distortion, then $D_C \leq D_{S_1}$. 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]. The paper by Goyal [5] provides a good overview of MDC systems.

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 [8] used MDC scalar quantizers to develop robust image and video coders for packet loss environments. Recently, Reibman [13] has, independently, proposed an MD video coder for packet networks 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); that is, if S_1 is the high-resolution (HR) description, then $D_{S_1} \leq D_{S_2}$ and $D_C = D_{S_1}$. 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 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. Moreover, though the UMDC encoder produces unbalanced descriptions, we use a 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 decoded. The main advantages of our MD video coder are stated as follows:

(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, that is, 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. 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, for example, [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 the one formed from past information produced by the central decoder; that is, 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 (Figure 1). Many MD video coders, for example, [16, 17, 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 (R-D) 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 in our system, in compliance with the standard syntax of H.263, an efficient end-of-block (EOB) symbol can be sent after the Kth 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 was 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 Recursive Optimal per-Pixel Estimate (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 intrablocks 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; that is, 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 this 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, 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, we 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 an 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, for example, [20], the signal is coded into a low rate base layer and a hierarchy of high rate enhancement layers. An enhancement layer is useful, that is, 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 UMDC system, the important layer is coded at 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 Figure 2. Two descriptions are generated, an HR (quality and bitrate) description and an LR 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 an HR block, that is, a block of the frame in the HR description, are duplicated 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.

The main contribution of this work is that the descriptions are generated in a rate-distortion optimal framework; that is, 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, we assume that there are N frames in the video sequence, with M macroblocks per frame, and let $d_{\mathrm{HR}}^{i,j}$ and $d_{\mathrm{LR}}^{i,j}$ represent the distortion in the jth macroblock

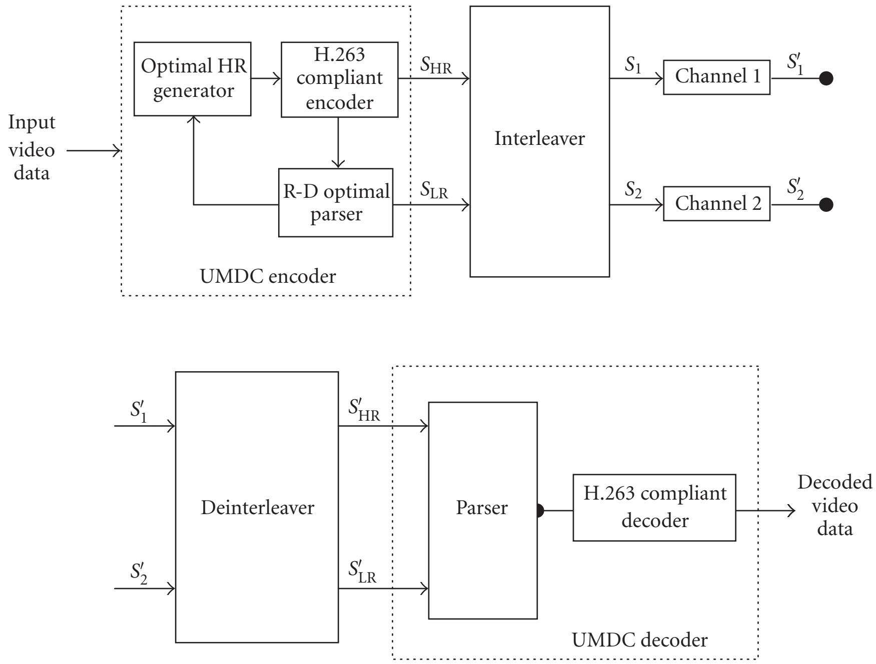


Figure 2: Proposed UMD coder block diagram. The input video sequence is coded into two descriptions; one description has high quality and bitrate and is obtained by coding the video sequence by an H.263 compliant encoder. The second description has low quality and bitrate and is obtained by selecting the important information from the HR 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.

of ith 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 the set of all macroblocks in the odd GOBs of a frame. Let r^i_{HR} and r^i_{LR} represent the ith frame rate for the HR and LR descriptions, respectively. Given the packetization policy, the expected distortion can be written as

$$\begin{aligned} E(D) = {} & \sum_i^N (1-p_i)^2 \cdot \sum_j^M d^{i,j}_{\mathrm{HR}} \\ & + \sum_i^N p_i \cdot (1-p_i) \cdot \left(\sum_{j \in E}^M d^{i,j}_{\mathrm{HR}} + \sum_{j \in O}^M d^{i,j}_{\mathrm{LR}} \right) \\ & + \sum_i^N p_i \cdot (1-p_1) \cdot \left(\sum_{j \in O}^M d^{i,j}_{\mathrm{HR}} + \sum_{j \in E}^M d^{i,j}_{\mathrm{LR}} \right), \end{aligned} \tag{1}$$

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

$$R_{\mathrm{HR}} + R_{\mathrm{LR}} = R_{\mathrm{TOT}}, \tag{2}$$

that is,

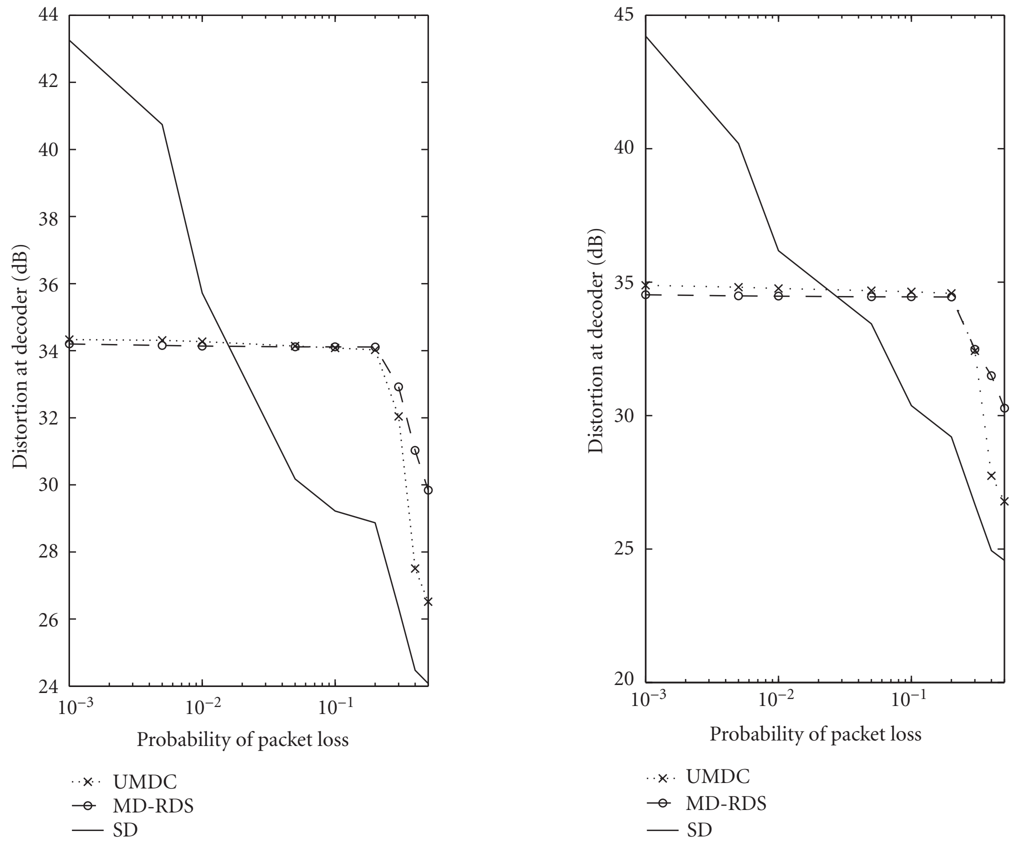
$$\sum_N r^i_{\mathrm{HR}} + \sum_N r^i_{\mathrm{LR}} = R_{\mathrm{TOT}}. \tag{3}$$

Solving this constrained optimization problem can be extremely complex. Due to the predictive nature of the video coder, $d^{i,j}_{\mathrm{HR}}$, r^j_{HR} depend on $d^{i-1,j}_{\mathrm{HR}}$, r^{i-1}_{HR} (this is true also for LR description). Further, in our MDC system, $d^{i,j}_{\mathrm{LR}}$ is also dependent on r^i_{HR} 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^i_{TOT} where

$$\begin{gathered} \sum_N r^i_{\mathrm{TOT}} = R_{\mathrm{TOT}}, \\ r^i_{\mathrm{HR}} + r^i_{\mathrm{LR}} = r^i_{\mathrm{TOT}}. \end{gathered} \tag{4}$$

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



(a) Akiyo. 256 kbps. 30 fps.

(b) Akiyo. 384 kbps. 30 fps.

Figure 3: UMDC versus MD-RDS and SD. Akiyo. 256 (a) and 384 (b) kbps.

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

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 the 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 jth 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, that is, not set to zero in *all* 8×8 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

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

where $\mathbf{Q}$ and $\mathbf{k}$ are the sets of quantization steps and the 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.

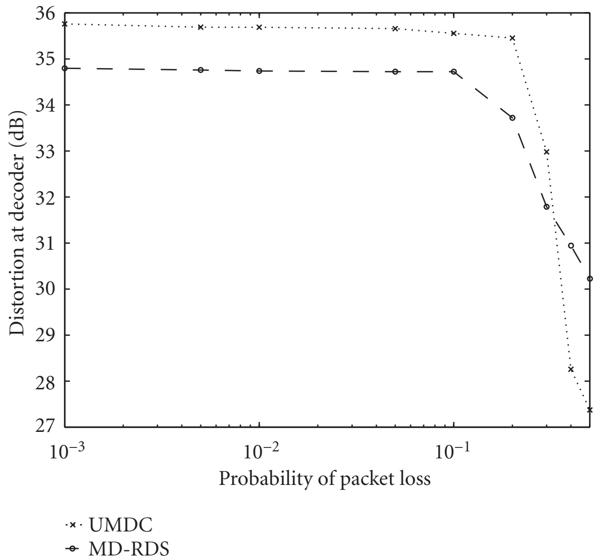


Figure 4: UMDC versus MD-RDS. Akiyo. 800 kbps. 30 fps.

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 ($l \in (1, \ldots, 4)$). Thus, $C^n_{l,j}$ is the nth coefficient in block l from the current macroblock j, and $\tilde{C}^n_{l,j}$ its quantized value, dependant on the quantization step (Q^j) used. We then can define the distortion in the HR and LR macroblocks as follows:

$$d^j_{\mathrm{HR}}(Q^j) = \sum_{l=1}^{4}\sum_{n=1}^{64}\left(C^n_{l,j} - \tilde{C}^n_{l,j}(Q^j)\right)^2, \tag{7}$$

$$d^j_{\mathrm{LR}}(Q^j, k^j) = \sum_{l=1}^{4}\left(\sum_{n=1}^{k^j}\left(C^n_{l,j} - \tilde{C}^n_{l,j}(Q^j)\right)^2 + \sum_{n=k^j+1}^{64}\left(C^n_{l,j}\right)^2\right). \tag{8}$$

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.

Using the TMN8 rate control algorithm find the required frame rate, r_{TOT}, for each frame. Then, for each frame of the input sequence,

(a) find all the $d_{\mathrm{HR}}(Q^j)$ and $d_{\mathrm{LR}}(Q^j, k^j)$ for all possible $\mathbf{Q}$ and $\mathbf{k}$;
(b) minimize the expected frame distortion (1) under the rate-budget r_{TOT} constraint for the frame in consideration (6);
(c) generate HR frame and LR frame according to the optimal $\mathbf{Q}$ and $\mathbf{k}$ found.

4. SIMULATIONS AND RESULTS

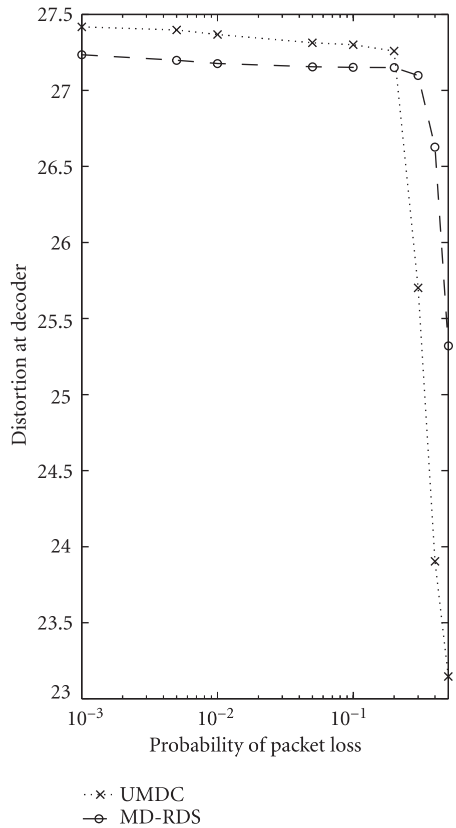
The reported results are expressed in terms of the signal-to-noise ratio (SNR) of the luminance components of the first 100 frames of the two QCIF (176×144 pixels) test sequences, namely, Akiyo and *Coastguard*. We compare the performance of our UMDC with the MD-RDS coder proposed by Reibman et al. [15]. 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 Section 3, 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.

Figures 3 and 4 show the results for the sequence Akiyo coded at different bitrates. The frame rate is 30 fps and only the initial frame of the stream is intracoded. 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 because MD-RDS alternates coefficients in order to make the descriptions balanced its coding efficiency decreases. Alternate nonzero coefficients adversely affect the entropy coder of standards like H.263. This implicit redundancy increases with the increase in the number of nonzero coefficients.

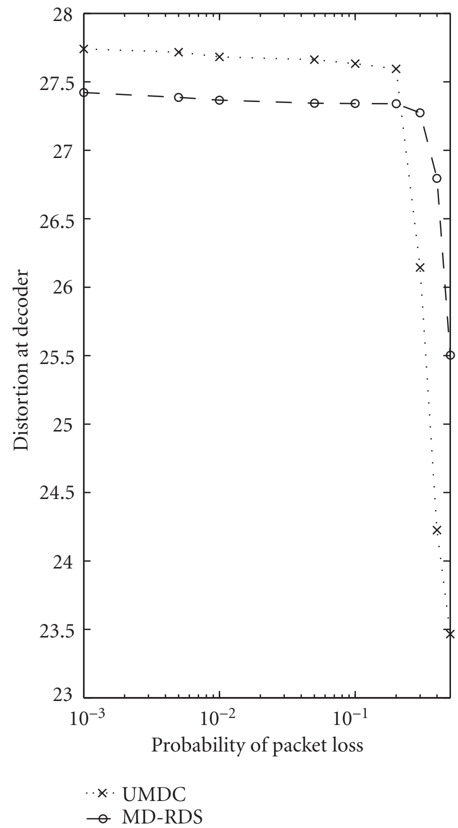
Plots in Figure 5 shows the results for the sequence *Coastguard*.

Figure 6 shows the comparison of both systems when only the first frame is coded in intraframe mode (Figure 6a) and when an intraframe is transmitted every 10 frames (Figure 6b). In this last case, a larger number of DCT coefficients are different from zero since intraframes contain a larger number of nonzero transform coefficients than interframes. Hence, the coding efficiency of the UMDC is much better than that of MD-RDS. The maximum gain over MD-RDS when an intraframe is inserted every 10 frames is around 1.3 dB, 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 single description. The test sequence used is Akiyo, coded at 512 kbps.

In Table 1, we compare UMDC with VRC. For VRC, we sent two packets per frame, where one packet contained



(a) Coastguard. 256 kbps. 30 fps.



(b) Coastguard. 384 kbps. 30 fps.

Figure 5: UMDC versus MD-RDS. Coastguard. 256 (a) and 384 (b) kbps.

Table 1: UMDC versus MD-RDS and versus VRC. Akiyo. 30 kbps. 10 fps.

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

only the even GOBs and the other packet odd GOBs. Akiyo sequence is coded at 30 kbps with frame rate 10 fps. Again UMDC does very well for low packet-loss scenarios.

5. CONCLUSIONS AND FUTURE WORK

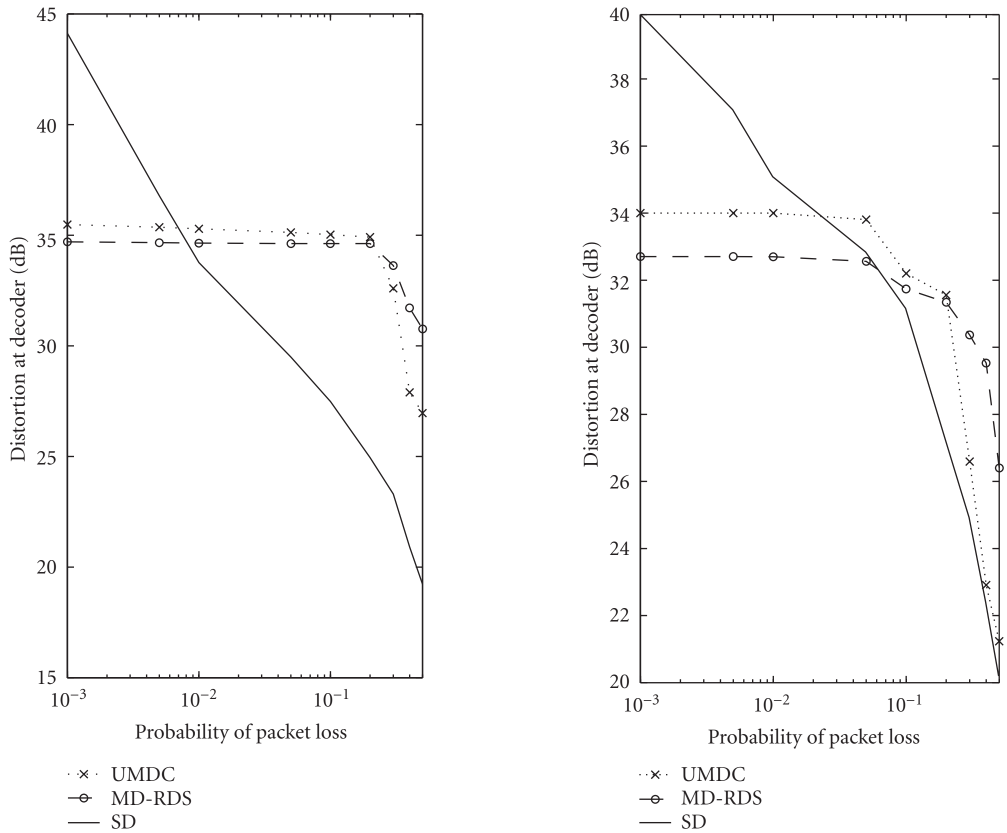
We have shown that the proposed UMDC system enhances the robustness of video coders to packet losses using a very small amount of extracomputational resources and being compliant with the existing video decoder standard syntax. The algorithm takes advantage of the processes done in the encoder and then, 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 UMDC 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.

In another future work we would like to use the LR sequence, not just as plain redundancy but also as a refiner of the HR description.



(a) Akiyo. 512 kbps. 1 Intra, 99 Inter.

(b) Akiyo. 512 kbps. 1 Intra every 10 frames.

Figure 6: UMDC versus MD-RDS and SD. Akiyo. 512 kbps. 30 fps.

ACKNOWLEDGMENTS

The authors would like to thank the reviewers for their helpful suggestions. This work was done while the first and the 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.

REFERENCES

[1] ITU-T Recommendation H.263, "Video Coding for low bit rate communication," International Telecommunications Union-Telecommunications, Geneva. 1, 1995; 2 (H.263+), 1998; 3 (H.263++), 2000.

[2] Motion Pictures Expert Group (ISO/IEC JTC 1/SC 2/WG 11), "Information Technology-Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to About 1.5 Mbits/s - Part 2: Video," Int. Std, 11172, November 1992.

[3] R. Rejaie, M. Handley, and D. Estrin, "Quality adaptation for congestion controlled video playback over the internet," in *Proc. ACM SIGCOMM '99*, pp. 189–200, Cambridge, Mass, USA, September 1999.

[4] S. Lin and D. J. Costello, *Error Control Coding: Fundamentals and Applications*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1983.

[5] V. K. Goyal, "Multiple description coding: Compression meets the network," *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp. 74–93, 2001.

[6] L. Ozarow, "On a source coding problem with two channels and three receivers," *Bell System Technical Journal*, vol. 59, no. 10, pp. 1909–1921, 1980.

[7] A. A. El Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Transactions on Information Theory*, vol. 28, no. 6, pp. 851–857, 1982.

[8] V. A. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 821–834, 1993.

[9] V. K. Goyal, J. Kovacevic, R. Arean, and M. Vetterli, "Multiple description transform coding of images," in *Proc. IEEE International Conference on Image Processing*, vol. 1, pp. 674–678, Chicago, Ill, USA, October 1998.

[10] Y. Wang, M. T. Orchard, and A. R. Reibman, "Optimal pairwise correlating transforms for multiple description coding," in *Proc. IEEE International Conference on Image Processing*, Chicago, Ill, USA, October 1998.

[11] R. Puri and K. Ramchandran, "Multiple description source coding through forward error correction codes," in *Proc. 33rd Asilomar Conference on Signals, Systems, and Computers*, vol. 1, pp. 342–346, October 1999.

[12] W. Jiang and A. Ortega, "Multiple description coding via polyphase transform and selective quantization," in *Proc. Visual Communications and Image Processing*, San Jose, Calif, USA, January 1999.

[13] A. R. Reibman, "Optimizing multiple description video coders in a packet loss environment," in *Proc. Packet Video Workshop*, Pittsburgh, Pa, USA, March 2002.

[14] R. Zhang, S. L. 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, pp. 966–976, 2000.

[15] A. R. Reibman, H. Jafarkhani, Y. Wang, and M. T. Orchard, "Multiple description video using rate-distortion splitting," in *Proc. IEEE International Conference on Image Processing*, Tessaloniki, Greece, October 2001.

[16] V. A. Vaishampayan and S. John, "Interframe balanced multiple description video compression," in *Proc. Packet Video Workshop*, New York, NY, USA, April 1999.

[17] A. R. Reibman, H. Jafarkhani, Y. Wang, M. T. Orchard, and R. Puri, "Multiple description coding for video using motion compensated prediction," in *Proc. IEEE International Conference on Image Processing*, Kobe, Japan, October 1999.

[18] Y. Wang and S. Lin, "Error resilient video coding using multiple description motion compensation," in *Proc. IEEE Multimedia Signal Processing Workshop*, Cannes, France, October 2001.

[19] S. Wenger, "Video redundancy coding in H.263+," in *Proc. Workshop on Audio-Visual Services for Packet Networks*, Aberdeen, UK, 1997.

[20] M. van der Schaar and H. Radha, "A hybrid temporal-SNR fine-granular scalability for internet video," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 11, no. 3, pp. 318–331, 2001.

[21] M. Gallant, G. Cote, B. Erol, and F. Kossentini, "UBC's H.263+ Public Domain Software, Version 3.2.0," Official ITU-T Study Group 16 Video Experts Group Reference C Codec, July 1998.

[22] A. Ortega and K. Ramchandran, "Rate-distortion methods for image and video compression," *IEEE Signal Processing Magazine*, vol. 15, no. 6, pp. 23–50, 1998.

[23] J. Ribas-Corbera and S. Lei, "Rate control in DCT video coding for low-delay communications," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 9, no. 1, pp. 172–185, 1999.

[24] D. Comas, R. Singh, and A. Ortega, "Rate-distortion optimization in a robust video transmission based on unbalanced multiple description coding," in *Proc. IEEE Workshop on Multimedia Signal Processing*, Cannes, France, October 2001.

David Comas was born in Barcelona (Spain) in 1976. In 2001, he received his M.S. degree in electrical engineering from the Technical University of Catalonia (UPC), Barcelona, Spain. From May 2000 to April 2001, he was with the Signal and Image Processing Institute (SIPI) at the University of Southern California (USC), Los Angeles, USA, doing his Master thesis. In 2001, he joined the Image Processing Group, in the Department of Signal Theory and Communications (TSC) and he is currently pursuing a Ph.D. degree in electrical engineering. His current research interests include multiple-description coding for video, image and sequence analysis, and image and sequence coding.

Raghavendra Singh was born in New Delhi, India, in 1971. He received the B.E. in electrical and electronics from the Birla Institute of Technology and Science, Pilani, India in 1993 and the Ph.D. in electrical engineering from University of Southern California in 2001. His research interests are in the area of digital image and video processing, and reliable transmission of real-time multimedia data.

Antonio Ortega received the Telecommunications Engineering degree from the Universidad Politecnica de Madrid (UPM), Madrid, Spain in 1989 and the Ph.D. in electrical engineering from Columbia University, New York, NY in 1994. At Columbia, he was supported by a Fulbright scholarship. In 1994, he joined the Electrical Engineering-System department at the University of Southern California, where he is currently an Associate Professor. At USC, he is also a member of the Integrated Media Systems Center (IMSC) and the Signal and Image Processing Institute. He is a senior member of the IEEE, and a member of SPIE and ACM. He was a Technical Program Cochair of ICME 2002. He has been an Associate Editor for the IEEE Transactions on Image Processing and for the IEEE Signal Processing Letters. He has been a member of the Board of Governors of the IEEE SPS. He has received the NSF CAREER Award, the 1997 IEEE Communications Society Leonard G. Abraham Prize Paper Award and the IEEE Signal Processing Society 1999 Magazine Award. His research interests are in the areas of multimedia compression and communications. They include topics such as rate control and video transmission over packet network and compression for recognition and classification applications.

Ferran Marqués received the M.S. degree in electrical engineering from the Technical University of Catalonia (UPC), Barcelona, Spain, in 1988. From 1989 to June 1990, he worked in the Swiss Federal Institute of Technology in Lausanne (EPFL) and in June 1990, he joined the Department of Signal Theory and Communications of the Technical University of Catalonia (UPC). In 1991, he was with the Signal and Image Processing Institute at USC in Los Angeles (California). He received the Ph.D. degree from the UPC in December 1992 and the Spanish Best Ph.D. thesis in electrical engineering award in 1992. Since 1995, he is an Associate Professor at UPC. He has served for the EURASIP AdCom as responsible of the member services (1998–2000), as Secretary and Treasurer (2000–2002), and currently as President. He was an Associate Editor of the Journal of Electronic Imaging (SPIE) in the area of image communications (1996–2000) and currently, he serves in the Editorial Board of the EURASIP Journal on Applied Signal Processing. He is an author or a coauthor of more than 70 publications that have appeared as journal papers and proceeding articles, 4 book chapters, and 4 international patents.