

# MULTIPLE DESCRIPTION WAVELET CODING OF LAYERED VIDEO

*N. V. Boulgouris<sup>1</sup>, K. E. Zachariadis<sup>2</sup>, A. Kanlis<sup>3</sup>, and M. G. Strintzis<sup>3</sup>*

<sup>1</sup>Electrical and Computer Engineering Department, University of Toronto, Canada  
<sup>2</sup>Electrical and Computer Engineering Department, Northwestern University, U.S.A.  
<sup>3</sup>Informatics and Telematics Institute, Thessaloniki 57001, Greece

## ABSTRACT

In this paper, a novel framework for multiple description coding of video is presented. The proposed scheme is based on a wavelet source coder and an efficient methodology for the generation of multiple descriptions. An algorithm is also presented for the optimal, in the sense of maximizing reconstruction quality, allocation of redundancy among the descriptions. Experimental results for the transmission of video using two descriptions demonstrate the efficiency of the proposed method.

## 1. INTRODUCTION

The transmission of pictorial information over today's heterogeneous, and often unreliable, networks (as in wireless applications) has necessitated the provision of coding methods which are able to deliver acceptable reconstruction quality even in case of channel failure. Multiple description (MD) coding investigates the problem of encoding a source into multiple independently decodable bitstreams which are mutually refining and equally important. The reconstruction quality at the decoder depends on the number of descriptions that are received. Due to its flexibility, multiple description coding is considered a very robust and reliable tool for information transmission.

Multiple Description coding has been recently studied for image and video transmission [1, 2, 3, 4]. The study of MD systems becomes more complicated in the case of video coding. The difficulty arises from the variety of different predictions that might be used at the decoder of an MD system. Several systems have been proposed for the MD coding of video.

In the present paper, the intra-frame and the motion compensated prediction residual frames are wavelet coded and divided into a redundant and an enhancement part with the redundant part encoded in all descriptions and the enhancement part distributed in several descriptions. The "repeat or split" strategy was chosen over other proposed techniques such as the one in [5] since, in our case, drift-free reconstruction is straightforward. Using the above framework, we present and evaluate two techniques for the multiple description coding of video sequences.

- In the first technique, *only the redundant part* is used for the construction of reference frames and thus, the resulting video coding scheme is able to perform drift-free reconstruction.
- In the second technique, *both the redundant and the non-redundant parts* of the stream are used for the creation of

the reference frame. This technique uses high-quality reference frames but the reconstructed video suffers from drift in case of transmission over channels with severe loss.

Additionally, in the present paper the problem of optimal redundancy allocation, i.e. appropriate selection of the redundant and the enhancement parts for each frame, is formulated as the maximization of the average video quality under the constraint of a target total rate.

## 2. MULTIPLE DESCRIPTION GENERATION

The proposed system for the generation of multiple descriptions is depicted in Figure 1. Initially, the available bit budget is evenly allocated to the frames in a GOP (Group of Pictures). The first frame in each GOP is intra-coded using block-based wavelet coding. The resulting coded stream is distributed over a number of descriptions. A portion of the bitstream is redundant in all descriptions. The correlation between consecutive frames is subsequently removed using Overlapped Block Motion Compensation (OBMC). The reference frames used to calculate motion vectors are the original frames in order to ensure good precision in the estimation of the motion vectors. Motion vectors are losslessly coded and are included in *all* descriptions.

Using the previously estimated half-pixel accurate motion vectors, the first inter-frame is compensated. We employ two different mechanisms for the derivation of reference frames that are used during motion compensation. In the first, a version of the I-frame, reconstructed using *only the redundant part* of the bitstream so far coded, is used as reference for the compensation process. In the second, *both redundant and non-redundant parts* are used for the derivation of reference frames in motion compensation. The prediction error is derived by subtracting the compensated prediction from the original inter-frame. The prediction error is wavelet transformed and coded into multiple descriptions. A version of the error frame is reconstructed using either the redundant part or both redundant and non redundant information of the coded bitstream depending on which of the two mechanisms described above is used. The reconstructed error frame is added to the compensated frame. The resulting inter-frame (instead of the original) will serve as the reference frame for the compensation of the next inter-frame. The same procedure is iterated until all frames in a GOP are treated.

The determination of the portion of the bitstream which is redundant in all descriptions is performed after the wavelet coding of the intra and the residual error frames. The wavelet coefficients are coded using a simple block-based bitplane encoder, based on the context models in [6]. Each block may be included in some or all descriptions. Thus, some blocks may appear in all descriptions

---

This work was completed while all authors were with the Informatics and Telematics Institute, Thessaloniki, Greece.

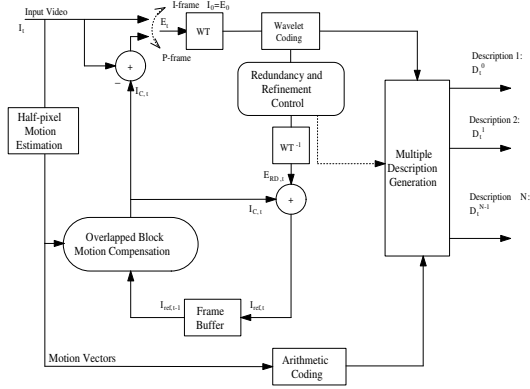


Fig. 1. Block diagram of the coder.

whereas some other blocks appear in only one of the descriptions. The inclusion of blocks in one or more descriptions is done so as to maximize the average quality at the decoder, subject to a total rate constraint, and attain fairly equal bitrate and fairly equal quality descriptions.

In order to apply an efficient redundancy allocation algorithm which takes into account the actual rate-distortion characteristics of the compressed stream, the distortion decrease achieved by the transmission of each bitplane should be calculated [6] for each layer. Each layer corresponding to a specific block of wavelet coefficients cause different reduction in the distortion. Let  $R_i$  be the number of bits required for the coding of the  $i$ th layer and  $D_i$  the corresponding distortion reduction they cause. When all pairs  $(D_i, R_i)$  are determined, the redundancy allocation algorithm can be applied.

An optimization algorithm should take into consideration the temporal correlation linking adjacent video frames. Modelling the dependency of adjacent frames in a video sequence is a nontrivial problem. In this paper, in order to deal with this issue, we assume that the total distortion reduction caused by the transmission of the  $i$ th layer in the  $m$ th frame can be expressed as

$$D_i + D_i C_i$$

where  $D_i C_i^1$  is the cumulative distortion reduction that is caused in the subsequent frames due to the higher quality current (reference) frame  $m$ . The coefficients  $C_i$  which quantify the impact of the current frame on the quality of subsequent frames were calculated using the methods in [7].

### 3. FORMULATION OF THE REDUNDANCY ALLOCATION PROBLEM

In order to address the problem of optimal allocation in MD video coding, it is important to derive expressions for the average video quality at the decoder and the total rate used in terms of the assignment strategy. Although in the experimental results section we consider the average PSNR over the entire sequence, in this section we shall attempt to maximize the distortion improvement incurred by each frame of the GOP separately. This simplification

<sup>1</sup>Even though coefficients  $D_i$  and  $C_i$  depend on the frame index  $m$ , this dependence will in the sequel be omitted for notational convenience.

will not significantly affect the optimality of the strategy derived here, while it will serve in addressing the problem of optimal assignment in a more rigorous way and in providing useful insight into the optimization procedure.

Let us assume that each frame is coded into  $L$  layers, each using  $R_i$  bits and contributing a reduction of distortion equal to  $D_i$  relative to the quality of the current frame and  $C_i D_i$ ,  $i = 1 \dots L$ , to the quality of the next frames in the GOP<sup>2</sup>, when used for motion compensation for the next frames. We further assume that the function  $D(R)$  of the distortion reduction with respect to the rate is concave. We shall consider the case of double description coding. The general case is studied in [8].

Let  $I = \{1, \dots, L\}$  denote the set of the possible values that the layer indices may assume. The problem of providing two descriptions for each frame in the GOP is equivalent to assigning a set of layer indices  $I_1 \subset I$  to the first and a set  $I_2 \subset I$  to the second description. Subsequently, the two descriptions will be transmitted over two communication links to the decoder. It can be shown [8] that the maximization problem can be formulated as the following:

**Maximization Problem** Find disjoint sets  $I_\cap, I_\cup \subset I$  maximizing

$$\bar{D}(I_\cap, I_\cup) = (2 - p) \sum_{i \in I_\cap} D_i(1 + C_i) + \sum_{i \in I_\cup} D_i \quad (1)$$

subject to the constraint

$$\bar{R}(I_\cap, I_\cup) = 2 \sum_{i \in I_\cap} R_i + \sum_{i \in I_\cup} R_i \leq R_B \quad (2)$$

where  $\bar{D}$  is the expected distortion,  $\bar{R}$  is the total rate,  $I_\cap \triangleq I_1 \cap I_2$  denotes the set of indices common to both descriptions, and  $I_\cup \triangleq (I_1 \cup I_2) \setminus I_\cap$  is the set of indices contained in exactly one of the descriptions.  $R_B$  denotes the total rate budget. The parameter  $p$  denotes the probability that each stream is successfully delivered to the decoder. The solution of the above problem will yield the optimal sets  $I_\cap$  and  $I_\cup$ , where  $I_\cap$  will contain the indices of the layers assigned to both streams and  $I_\cup$  will contain the indices assigned only to one of the streams. In order to obtain the optimal  $I_1, I_2$ , we need to further partition  $I_\cup$  into two disjoint index-assignment sets, one for each stream. It is clear however, that any such partition will yield sets  $I_1, I_2$  inducing the same expected distortion improvement at the decoder; hence, the partition of  $I_\cup$  may be arbitrary (we may even assign the whole set  $I_\cup$  to only one of the streams). However, since balanced MD coding is sought, an acceptable partitioning should result in fairly equal total rates of  $I_1$  and  $I_2$ . In order to achieve this, the indices in  $I_\cup$  may be ordered in terms of decreasing corresponding rates  $R_i$  and be assigned alternately to each stream.

### 4. FAST REDUNDANCY ALLOCATION

If we were to solve the maximization problem by exhaustively examining all possible realizations of  $I_1$  and  $I_2$ , this would involve  $2^{2L}$  possibilities, since there are  $2^L$  subsets of the index set  $I$ . Clearly, the optimal solution will be achieved by choosing any pair of sets  $I_1$  and  $I_2$  resulting in the same sets  $I_\cap^*$  and  $I_\cup^*$ , which solve the maximization problem described by (1) and (2). Hence, we only need to examine all possible realizations of disjoint sets  $I_\cap, I_\cup \subset I$ .

<sup>2</sup>for the last frame in the GOP  $C_i = 0, i = 1, \dots, L$ .

Note that since there are  $2^L$  possible subsets of the index set  $I$ , any subset  $A \subset I$  may be expressed as the binary representation of a number between 0 and  $2^L - 1$ , with the  $i^{th}$  bit being 1, if  $i \in A$  and 0 otherwise. An exhaustive search algorithm could determine the optimal solution  $I_\cap^*, I_\cup^*$  to the maximization problem. Although such an algorithm will always produce an optimal solution, the number of possible realizations of  $I_\cap$  and  $I_\cup$ , over which the search will be performed, is  $3^L$ , still prohibitive even for moderate values of  $L$ .

In order to reach a fast solution, we formulate a continuous maximization problem. By examining closely the Discrete Maximization Problem described by (1) and (2), we first note that the sums  $\sum_{i \in I_\cap} D_i(1 + C_i)$ ,  $\sum_{i \in I_\cap} R_i$  and  $\sum_{i \in I_\cup} D_i$ ,  $\sum_{i \in I_\cup} R_i$  are the distortion-improvement and rate "measures" of  $I_\cap$  and  $I_\cup$  respectively. A further restriction arises from the requirement that  $I_\cap$  and  $I_\cup$  have to comprise intervals dictated by the available blocks and that partial blocks may not be used. If we relax this restriction, we may formulate a corresponding Continuous Maximization Problem, which is easier to solve.

Assume that  $D(R)$  represents a continuous, differentiable, non-decreasing and concave function of the rate  $R$ . Then the derivative  $D'(R)$  will be a well-defined, continuous, positive and decreasing function of  $R$ , for every  $R \in \mathbb{R}_+$ . In a similar fashion, assume that the fraction of distortion improvement due to motion compensation is provided by a continuous decreasing function  $c(R)$  and that the curve corresponding to the products  $D_i C_i$  defines a function  $C(R)$  with derivative  $C'(R) = D'(R)c(R)$ , which will have properties similar to those of  $D'(R)$ <sup>3</sup>.

It can be shown [8] that an Equivalent Continuous Problem may be stated as follows:

**Equivalent Continuous Maximization Problem** Find positive rates  $R_\cup \geq R_\cap \geq 0$  maximizing

$$\bar{D}(R_\cap, R_\cup) = (2-p)[C(R_\cap) + D(R_\cap)] + D(R_\cup) - D(R_\cap)$$

subject to the constraint

$$\bar{R}(R_\cap, R_\cup) = R_\cap + R_\cup \leq R_B \quad (3)$$

This is a simple Lagrangian maximization problem with optimal solution  $R_\cap^*, R_\cup^*$  satisfying the constraint (3) at the boundary. The optimal  $R_\cap^*$  should satisfy

$$(2-p)[C'(R_\cap^*) + D'(R_\cap^*)] - D'(R_B - R_\cap^*) - D'(R_\cap^*) = 0,$$

which, after some simple manipulations translates to the condition

$$\phi(R_\cap^*) \triangleq \frac{D'(R_B - R_\cap^*)}{D'(R_\cap^*)} - (2-p)c(R_\cap^*) = 1-p. \quad (4)$$

Observe that, since  $D'(\cdot)$  and  $c(\cdot)$  are decreasing,  $\phi(\cdot)$  will be continuous and increasing in the interval  $[0, R_B/2]$  and the continuous maximization problem will not involve local maxima. Also, the smallest value of  $\phi(\cdot)$  will be  $\phi(0) = D'(R_B)/D'(0) - (2-p)c(0)$  and the largest value will be  $\phi(R_B/2) = 1 - (2-p)c(R_B/2)$ . Therefore, if  $(1-p) \in [\phi(0), \phi(R_B/2)]$ , the optimal value for  $R_\cap^*$  will be  $\phi^{-1}(1-p)$ . Otherwise (4) does not have a solution and optimality is achieved either at 0 or  $R_B/2$ . In general, we can write

$$R_\cap^* = \begin{cases} 0, & \text{if } 1-p < \phi(0) \\ \phi^{-1}(1-p), & \text{if } 1-p \in [\phi(0), \phi(R_B/2)] \\ R_B/2, & \text{if } 1-p > \phi(R_B/2) \end{cases} \quad (5)$$

<sup>3</sup>In other words,  $D'(R)$  corresponds to the ratios  $D_i/R_i$  and  $c(R)$  to the coefficients  $C_i$ .

while  $R_\cup^* = R_B - R_\cap^*$ .

Returning to the Discrete Maximization Problem, it is reasonable to assume that a near-optimal solution will resemble that of the Equivalent Continuous Maximization Problem, especially for large values of  $L$ . This means that a near-optimal choice for the index assignment sets would be  $I_\cap = \{1, \dots, L_\cap^*\}$ ,  $I_\cup = \{L_\cap^* + 1, \dots, L_\cup^*\}$ , where  $L_\cap^*$  and  $L_\cup^*$  would be such that

$$\sum_{i=1}^{L_\cap^*} R_i \simeq R_B - \sum_{i=1}^{L_\cup^*} R_i$$

and

$$\phi(L_\cap^*, L_\cup^*) \triangleq \frac{D_{L_\cup^*}/R_{L_\cup^*}}{D_{L_\cap^*}/R_{L_\cap^*}} - (2-p)c_{L_\cap^*} \simeq 1-p, \quad (6)$$

This consideration suggests the following

**Algorithm 1** (Fast Search Algorithm)

$L_1 = \max\{l \in I : \sum_{i=1}^l R_i \leq R_B/2\}$  (index corresponding to  $R_B/2$ )

$L_2 = \max\{l \in I : \sum_{i=1}^l R_i \leq R_B\}$  (index corresponding to  $R_B$ )

if  $\phi(L_1, L_2) > 1-p$ , set  $L_\cap^* = 0$ ,  $L_\cup^* = L_2$  and exit (case  $1-p < \phi(0)$  in (5))

if  $\phi(L_1, L_1) < 1-p$ , set  $L_\cap^* = L_1$ ,  $L_\cup^* = L_1$  and exit (case  $1-p > \phi(R_B/2)$  in (5))

$L_\cap = 1$ ,  $L_\cup = L_2$

(initial values)

while  $L_\cap \leq L_\cup$

while  $\sum_{i=1}^{L_\cup} R_i > R_B - \sum_{i=1}^{L_\cap} R_i$  decrease  $L_\cup$  (find largest  $L_\cup$  satisfying rate constraint)

if  $\phi(L_\cap, L_\cup) > 1-p$  set  $L_\cap^* = L_\cap$ ,  $L_\cup^* = L_\cup$  and exit (crossed  $1-p$  line)

increase  $L_\cap$

(next value of  $L_\cap$ )

endwhile

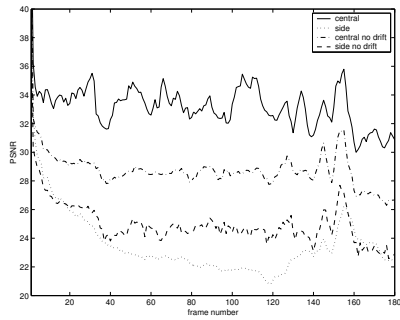
$L_\cap^* = L_1$ ,  $L_\cup^* = L_1$  (if this point is reached,  $\phi(L_\cap, L_\cup)$  never crossed the  $1-p$  line)

The advantage of this algorithm lies in that it involves fewer calculations and terminates far sooner than the exhaustive search algorithm. It is clear, however, that the price paid for its reduced complexity, which is important in cases of real-time applications, is its inferior performance compared to the exhaustive search algorithms.

## 5. EXPERIMENTAL RESULTS

The proposed Multiple Description video coding scheme was experimentally evaluated for the transmission of the Y component (15 frames/second) of the standard test sequence "Foreman" over two channels. Each frame was coded in two descriptions. Motion vector information was duplicated in both descriptions. The proposed Fast search redundancy allocation algorithm of the preceding section was applied for video transmission over two channels of total capacity 128 Kbps and for three different probabilities of description arrival:  $p = 0.8, 0.9, 0.95$ , or equivalently three probabilities of description loss equal to 20%, 10%, 5%. The number of frames in each GOP was chosen with respect to  $p$ . The target rate  $R_B$  for each frame was determined by allocating to Intra-frames rate equal to four times the rate allocated to Inter-frames. Simulations indicated that the resulting descriptions are remarkably "balanced" i.e. they have approximately equal size and yield almost equal reconstruction qualities.

We examine two types of transmission scenarios: in the first scenario, we assume that the channels retain their status during the entire transmission. In this case, the parameter  $p$  serves as a means to control the redundancy and is not directly associated with the condition of the channel. In the second scenario, we assume that the channels go on and off during transmission. In the latter scenario, it is possible that both descriptions of a frame are lost. In such a case the decoder uses the most recent reference frame that is available. For each frame, the Peak-Signal-to-Noise-Ratio is used as a measure of the reconstruction quality (in dB).

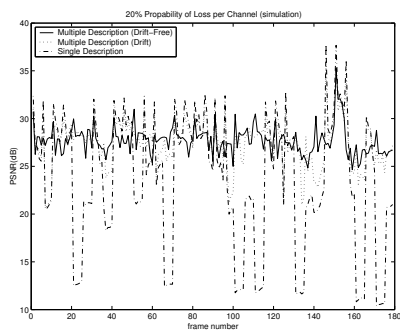


**Fig. 2.** Reconstruction quality for the “Foreman” sequence using the present coders.

In Fig. 2, the coding of the “Foreman” sequence into two descriptions is simulated under the respective assumption that the channels are available or unavailable during the entire transmission. As expected, the central distortion in the proposed scheme that allows drift accumulation is superior in comparison to the proposed drift-free system. The side distortion, however, appears to be better in the drift-free system.

Packet Loss	Single	Drift	Drift-Free
20%	17.95	26.60	26.90
10%	21.70	27.32	27.49
5%	22.96	27.60	27.84

**Table 1.** Performance comparison. PSNR (in dB) is reported.



**Fig. 3.** Reconstruction quality for the “Foreman” sequence when the channels go on and off during transmission.

In the second simulation, in which the channels may go on and off from frame to frame, we tested our systems under identical description loss patterns. For each frame, one, two or none of the descriptions were lost. As seen from Table 1 and Fig. 3, the drift-free system is much more reliable and demonstrates no abrupt changes in its performance, contrary to the drift-allowing scheme which demonstrates significant variations in the video quality it delivers. In addition, both schemes demonstrate significant gains over the single description scheme which appears to collapse very frequently due to description losses.

## 6. CONCLUSIONS

In this paper, a novel framework for multiple description coding of video was presented. The underlying scalable video coder uses motion compensation and wavelet decorrelation of compensation residuals. The coded stream is organized into multiple descriptions in such a way that redundancy is efficiently exploited to eliminate drift at the decoder. An algorithm for the allocation of redundancy among the descriptions was also proposed. Experimental results using the drift-free framework and the redundancy allocation algorithm demonstrated the efficiency of the proposed methodology.

## 7. REFERENCES

- [1] A. R. Reibman, H. Jafarkhani, Y. Wang, M. T. Orchard, and R. Puri, “Multiple Description Coding for Video using Motion Compensated Prediction,” in *Proc. ICIP 1999*, Kobe, Japan, October 1999.
- [2] V. Vaishampayan and S. John, “Interframe Balanced Multiple Description Video Compression,” in *Proc. Packet Video 1999*, New York, USA, April 1999.
- [3] N. V. Boulgouris, K. E. Zachariadis, A. Leontaris, and M. G. Strintzis, “Drift-free multiple description coding of video,” in *Proc. IEEE Int. Symposium on Multimedia Signal Processing*, Cannes, France, Oct. 2001.
- [4] X. Tang and A. Zakhor, “Matching Pursuit Multiple Description Coding for Wireless Video,” *IEEE Trans. Circuits and Systems for Video Technology*, vol. 12, no. 6, pp. 566–575, June 2002.
- [5] Y. Wang, M. T. Orchard, V. Vaishampayan, and A. R. Reibman, “Multiple Description Coding Using Pairwise Correlating Transforms,” *IEEE Trans. Image Processing*, vol. 10, no. 3, March 2001.
- [6] D. Taubman, “High performance scalable image compression with EBCOT,” *IEEE Trans. Image Processing*, vol. 9, no. 7, pp. 1158–1170, July 2000.
- [7] P.-Y. Cheng, J. Li, and C.-C. J. Kuo, “Rate control for an embedded wavelet video coder,” *IEEE Trans. Circuits and Systems for Video Technology*, vol. 7, no. 4, pp. 696–702, 1997.
- [8] N. V. Boulgouris, K. E. Zachariadis, A. Kanlis, and M. G. Strintzis, “Multiple description wavelet coding of layered video using efficient redundancy allocation,” *IEEE Trans. Circuits and Systems for Video Technology*, submitted for publication.