A Flexible R-D-Based Multiple Description Scheme for JPEG 2000

Enrico Baccaglini, *Student Member, IEEE*, Tammam Tillo, *Member, IEEE*, and Gabriella Olmo, *Member, IEEE*

***Abstract*—Multiple description coding (MDC) is a good way to combat packet losses in error-prone networks subject to packet erasures. However, redundancy tuning is often difficult, and this makes the generation of descriptions with good redundancy-rate-distortion performance a hard job. Moreover, the complexity of generating more than two descriptions represents a strong limitation to MDC. In this letter, we propose a method that exploits the data rate-distortion characteristics to generate multiple descriptions for JPEG 2000 with tunable redundancy levels. The identification of the best number of descriptions as a function of the network conditions is also addressed.**

***Index Terms*—Image coding, image communication, multiple description coding (MDC), source channel coding.**

I. Introduction

In modern wireless multimedia communications, a common scenario encompasses many receivers scattered over a heterogenous network and equipped with different capabilities in terms of memory, visual resolution, and computational resources. Moreover, different virtual paths are available for data to reach each receiver, and physical links between nodes are unreliable and subject to bit errors and independent packet losses.

This scenario strongly calls for coding techniques where all the received packets can be exploited at the application layer. It is well known that multiple description coding (MDC) is an effective method to protect multimedia information transmitted over nonprioritized networks [3]. In the MDC approaches, two or more nonhierarchical representations of the same data are generated, yielding mutually refinable information. The number of descriptions and the introduced redundancy should be carefully tuned in order to match the network conditions.

The most popular MDC methods stem from the pioneering MD scalar quantization (MDSQ) [4]. In the two-description case, the index information of a standard quantizer is arranged on the diagonals of a matrix. The matrix structure and the number of diagonals filled with indexes affect the tuning of redundancy. In the case of more than two descriptions, controlling the redundancy and generating descriptions becomes a complicated job [5]. Another class of methods employs pairwise correlating transforms operating on the coefficients in order to introduce a controlled amount of redundancy among the descriptions [7]; however, for more than two descriptions, grouping coefficients and allocating redundancy is not trivial [8]. Methods based on the polyphase decomposition followed by selective quantization [9] and rate allocation [10] are more flexible, making such algorithms a good choice for applications where the probability of description loss widely changes, such as wireless applications. The results reported in [10] demonstrate the superior performance of a rate allocation scheme based on two possible rates over other state-of-the-art MD methods for JPEG 2000. In [1], an analytical bound has been worked out that determines when a two-description configuration outperforms (in terms of perceived quality) the standard single-description coding. In [2], a method for generating descriptions starting from more than two encoding rates has been proposed for a Gaussian source.

In this letter, we propose an MD scheme for JPEG 2000 that, stemming from the results in [1] and [10], allows one to find the analytical bounds for an arbitrary number of descriptions and two different encoding rates. The proposed procedure yields the optimal number of descriptions and the best redundancy level based on the network loss probability and the data rate-distortion (R-D) curve. This allows to achieve good performance over a wide range of packet loss rates.

II. Proposed Algorithm for Generating Balanced Descriptions

In order to build N balanced descriptions[1] for a given image, we generate two different JPEG 2000 bitstreams of rates R_1 bits per pixel (bpp) and $R_2 < R_1$ bpp.

The codeblocks (CBs) generated after the wavelet decomposition of the JPEG 2000 encoding procedure [11] are grouped into N sets. These sets are built so that they have similar characteristics in terms of size and distortion contribution. In principle, to obtain N balanced descriptions, one should identify subsets of CBs with similar R-D characteristics and allocate them to the descriptions. Nevertheless, we have verified that, if the number of CBs per subband is at least equal to N, it is sufficient that each description contains the same number of CBs encoded at rate R_1 (and consequently at rate R_2) for each subband.

The first description is obtained by combining the CBs of set 1 taken from the first stream, with the CBs of all the other sets taken from the second one. Analogously, the ith description is built by taking the CBs of set i from the first stream and those of the remaining sets from the second stream.

This procedure generates an arbitrary number N of balanced descriptions fully backward compatible with Part 1 of the JPEG

Manuscript received May 16, 2006; revised August 7, 2006. This work was supported by the European Union under project NEWCOM, "Network of Excellence in Wireless Communications." The associate editor coordinating the review of this manuscript and approving it for publication was Prof. James E. Fowler.

The authors are with CERCOM, Center for Multimedia Radio Communications, Politecnico di Torino, 10129 Torino, Italy (e-mail: enrico.baccaglini@polito.it; tammam.tillo@polito.it; gabriella.olmo@polito.it).

Digital Object Identifier 10.1109/LSP.2006.883986

[1] Two equal rate descriptions are said to be balanced if they result in identical average distortion when used individually [4].

2000 standard [10], each of which is encoded at $((R_1 + (N-1)R_2)/N)$ bpp, yielding a total output rate of $R = R_1 + (N-1)R_2$ bpp. In the rest of this letter, we assume that R is fixed by the application requirements and kept constant during the data transmission.

The decoder preprocesses all the received descriptions and selects, for each CB, the finest representation. Given that the procedure yields balanced descriptions, when a subset of k descriptions out of N is received, a side distortion $D_{k,N}$ is obtained, representing a weighted average of the quality yielded by two bitstreams at rates R_1 and R_2. As the distortion of JPEG 2000 CBs can be assumed additive [11], we can write that

$$D_{k,N}(R_2) = \frac{(N-k)D(R_2) + kD(R-(N-1)R_2)}{N}, \quad 1 \le k \le N \quad (1)$$

where $D(r)$ is the R-D curve of the JPEG2000 image encoded at rate r. When all descriptions are received, the quality is a function of R_1 only, because all CBs are taken from the bitstream coded at that rate. Therefore, the central distortion can be evaluated as $D(R_1) = D(R-(N-1)R_2)$. The extra rate, which amounts to $(N-1)R_2$, is referred to as the *redundancy* of the MDC scheme. Due to the bound on the overall rate R, when all descriptions are received, the redundancy impairs the R-D performance of the system, but it becomes beneficial when a subset of descriptions is received. As a consequence, the redundancy must be carefully tuned according to the network conditions, and this can be accomplished by modifying the fraction of total rate to be assigned to the stream at lower rate R_2.

III. Theoretical Analysis

A. Two-Description Case

In the two-description case, the previously described procedure yields descriptions encoded at rate $(R_1 + R_2)/2$ bpp. According to (1), the side distortion is

$$D_{1,2}(R_2) = \frac{D(R_2) + D(R-R_2)}{2}. \quad (2)$$

Assuming that each description fits a single packet, packets are randomly dropped, and no Automatic Repeat reQuest (ARQ) is used, we can work out the expected end-to-end distortion for two descriptions

$$\overline{D}_2(R_2) = (1-p)^2 D(R-R_2) + 2p(1-p)D_{1,2}(R_2) + p^2\sigma^2 \quad (3)$$

where p is the probability of packet loss,[2] and σ^2 is the variance of the image. Using (2), we can write that

$$\overline{D}_2(R_2) = (1-p)\left[D(R-R_2) + pD(R_2)\right] + p^2\sigma^2. \quad (4)$$

The MDC scheme should be optimized selecting the value of redundancy R_2 that minimizes (4), given p and the total rate R.

[2]If a description does not fit a single packet, the probability p should be replaced by the probability of description loss. The relationship between probability of packet loss and probability of description loss depends on the actual packetization scheme and description size.

This optimization task will be addressed in Section III-B in the general case of N descriptions.

In order to make sensible evaluations of the effectiveness of the proposed two-description scheme, we compare it with a single-description coding (SDC) scheme where the source is encoded at rate $R = R_1 + R_2$ bpp and fit into two packets.

When the source image is single-description encoded, the expected end-to-end distortion can be evaluated as

$$\overline{D}_1 = (1-p)^2 D(R) + p(1-p)D_{pkt_1} + p(1-p)D_{pkt_2} + p^2\sigma^2 \quad (5)$$

where D_{pkt_i} is the distortion when only packet i ($i = 1, 2$) is received and $\sigma^2 \ge D_{pkt_i} \ge D(R)$, $i = 1, 2$. If only one of the two packets (e.g., pkt_i) can be decoded independently of the other (e.g., pkt_j), then $D_{pkt_i} \ge D(R)$ and $D_{pkt_j} = \sigma^2$. This is a typical situation of layered coding, when the information contained in packet j is usable if and only if packet i has been properly received. In this situation, (5) is lower-bounded by $\overline{D}_{LB}$

$$\overline{D}_1 \ge (1-p)D(R) + p\sigma^2 = \overline{D}_{LB} \quad (6)$$

where $\overline{D}_{LB}$ also represents the expected end-to-end distortion of the SDC scheme when all data fit a single packet and assuming that p is almost independent of the packet length (as inferred in [6]).

If both packets need to be received in order to decode the source, then $D_{pkt_i} = D_{pkt_j} = \sigma^2$, and the lower bound (6) still holds. If both packets can be independently decoded, then the system turns out to be an MDC one.

In order to determine if MDC outperforms SDC, according to (4) and (6), we work out the condition under which $\overline{D}_{LB} - \overline{D}_2 > 0$, obtaining

$$\frac{D(R-R_2) - D(R)}{\sigma^2 - D(R_2)} = f(R_2) < p. \quad (7)$$

The optimal solution is then obtained selecting R_2 so that $f(R_2)$ is minimized. If it happens that $p_2 = \min_{R_2} f(R_2) \ge p$, then the SDC scheme always outperforms MDC.

From inequality (7), we infer that, in order to decide whether it is convenient to transmit data as one or two descriptions, it is sufficient to estimate p and the R-D function of the data. The probability p can be estimated or obtained using a feedback channel.

B. Arbitrary Number of Descriptions

In this section, we study the general case of an arbitrary number N of descriptions generated using the procedure of Section II. We identify the optimal number of descriptions and the related redundancy according to the network conditions and the available resources. As stated in Section III-A, assuming that each description fits a packet and that no ARQ is employed, we can work out the following expression of the expected end-to-end distortion:

$$\overline{D}_N(R_2) = \sum_{k=1}^{N} \binom{N}{k} (1-p)^k p^{N-k} D_{k,N}(R_2) + p^N\sigma^2 \quad (8)$$

where $D_{k,N}(R_2)$ is reported in (1). Taking into account that

$$\sum_{k=1}^{N}\binom{N}{k}\frac{N-k}{N}(1-p)^k p^{N-k} = p - p^N$$
$$\sum_{k=1}^{N}\binom{N}{k}\frac{k}{N}(1-p)^k p^{N-k} = 1 - p$$

we can rewrite (8) as

$$\overline{D}_N(R_2) = (p - p^N)D(R_2) + (1-p)D(R_1) + p^N\sigma^2 \quad (9)$$

where $R_1 = R - (N-1)R_2$. We want to find out how to optimally split the total available rate R between R_1 and R_2. To this end, we work out the minimum of (9), given N, p, and R. Evaluating the derivative of (9), with respect to R_2, we obtain that

$$\frac{\partial \overline{D}_N(R_2)}{\partial R_2} = (p-p^N)\frac{\partial D(R_2)}{\partial R_2} + (1-p)\frac{\partial D(R_1)}{\partial R_1}\frac{\partial R_1}{\partial R_2}. \quad (10)$$

The minimum expected end-to-end distortion $(\overline{D}_N)_{min}$ is then obtained from (9) selecting R_2 so that the following relationship is satisfied:

$$\frac{\frac{\partial D(R_2)}{\partial R_2}}{\frac{\partial D(R_1)}{\partial R_1}} = \frac{(1-p)(N-1)}{p-p^N}. \quad (11)$$

In order to find the optimal number of descriptions N_{opt} from a given set of candidate values, an exhaustive search may be feasible if the set is not too large. In fact, using (9) and (11), the optimal rates R_1, R_2, and the minimum possible distortion for that N can be analytically worked out without the need of actual co-decoding. In case exhaustive search cannot be afforded, the greedy approach sketched below has shown to yield satisfactory performance:

1) Set $N = 2$.
2) Initialize M to a candidate value larger than N.
3) $(\overline{D}_N)_{min} \le (\overline{D}_M)_{min}$?
    3a) NO: set $N = M$; assign a larger candidate value to M and go to 3.
    3b) YES: $N_{opt} = N$; end.

IV. Experimental Results and Conclusions

In [10], the two-description case of the algorithm has been compared with state-of-the-art procedures for JPEG2000. We now focus on the possibility of tuning redundancy and number of descriptions. The algorithm has been tested using the JPEG 2000 Verification Model 8.6 co-decoder and the test images Lenna, Goldhill, and Elaine of dimension 512 × 512 pixels. The wavelet transform has been applied with four levels of decomposition, and for each image, the total output rate has been set to $R = 0.4$ and $R = 1.2$ bpp. SDC and MDC with $N = 2$, 4, and 8 descriptions have been compared. Each description fits a packet as discussed in Section III. The SDC performance bound is evaluated according to (6).

In Fig. 1, the end-to-end PSNR is reported versus p for image Lenna. The curves have been worked out by simulation, given the total rate R, and selecting the best R_2 for each value of p. It can be noticed that, as p increases, a higher number N of descriptions leads to higher end-to-end PSNR. For example, at 0.4 bpp and $0.06 \le p \le 0.36$, the encoder should create four descriptions to ensure a gain up to 4 dB over the two-description configuration and up to 9 dB if compared to SDC. Then, for $p > 0.36$, the eight-description scheme outperforms those with a smaller number of descriptions. In fact, many descriptions allow the transmission to rely on many representations of the source, and the reception of a subset of them can be sufficient to reconstruct the original information with decent quality.

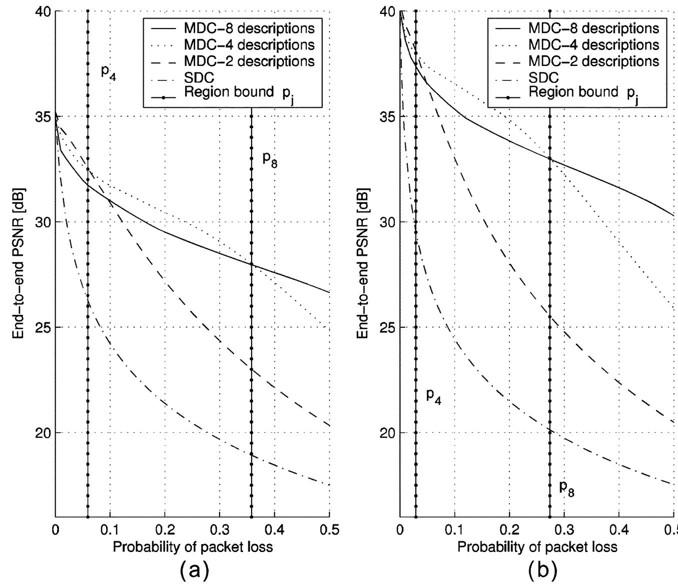

Fig. 1. Expected end-to-end PSNR versus probability p of packet loss for image Lenna. (a) $R = 0.4$ bpp. (b) $R = 1.2$ bpp.

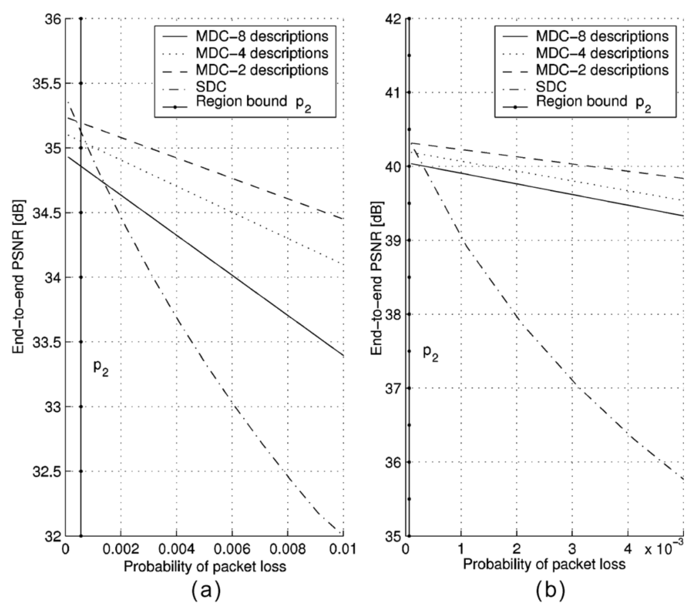

Fig. 2. Expected end-to-end PSNR versus low values of probability of packet loss p for image Lenna. (a) $R = 0.4$ bpp. (b) $R = 1.2$ bpp.

TABLE I
SIMULATION AND ANALYTICAL BOUNDS FOR IMAGES LENNA, GOLDHILL, AND ELAINE AT RATES 0.4 AND 1.2 bpp

Image	R	p_2		p_4		p_8	
		Simulation	Analytic	Simulation	Analytic	Simulation	Analytic
Lenna	0.4	4×10^{-4}	5.5×10^{-4}	0.063	0.059	0.36	0.36
	1.2	1×10^{-4}	0.7×10^{-4}	0.032	0.029	0.27	0.27
Goldhill	0.4	0.8×10^{-3}	1.6×10^{-3}	0.051	0.057	0.33	0.33
	1.2	2.5×10^{-4}	2.5×10^{-4}	0.032	0.036	0.27	0.29
Elaine	0.4	3.5×10^{-4}	4.5×10^{-4}	0.049	0.043	0.36	0.32
	1.2	3×10^{-4}	2×10^{-4}	0.032	0.032	0.26	0.24

It can also be noticed that the scheme gets inefficient in the high-redundancy region. This is due to the fact that most information is duplicated in this case. A possible solution to improve the algorithm performance could be the use of interleaved quantizers (MDSQ-like [4]), but as a consequence, the descriptions would not be backward compatible with Part 1 of the JPEG 2000 any more. Fig. 1 also reports the bounds p_j that delimit the regions where j outperform i descriptions, with $i < j$. Such boundary values [with possible exception of p_2, which can be analytically worked out using (7)] are obtained selecting the intersections between curves $\overline{D}_i(R_2)$ and $\overline{D}_j(R_2)$ of (9) with R_2 optimized using (11).

Fig. 2 focuses on low values of p, comparing SDC and MDC. The probability p_2 derived from (7) is also reported, which represents the analytical lower bound of region where MDC is advantageous with respect to SDC. It can be noticed that the values of p for which SDC outperforms MDC are extremely low. This is a consequence of the fact that, in our scheme, the redundancy can be finely tuned; this allows for an almost continuous transition between the single and multiple description cases. The SDC is nearly included in MDC scheme in the limit $R_2 \rightarrow 0$.

Table I reports, for the three test images, region bounds and intersections obtained from simulation. A very good match between such boundaries can be appreciated; the difference is mainly due to the approximations of the level of redundancy in the simulation chain. This implies that, once the probability of packet loss has been estimated, it is possible to recursively run the algorithm of Section III varying the parameter N to get its optimal value in order to dynamically adapt the encoder to the network conditions.

Finally, we tested the performance sensitivity to estimated packet loss probability; in fact, this parameter may be difficult to precisely estimate, or it may change dynamically. Therefore, it is important to verify the effects of a possible mismatch between the actual p_{act} and the estimated value p. Results are reported in Table II. We analyzed the transmission of images Lenna and Goldhill at 1.2 bpp using two descriptions and evaluated the average quality (in terms of PSNR) when receiving one or two descriptions. As expected, if the actual loss rate p_{act} matches p, the system yields the best results compared to other redundancy allocations. If p is overestimated with respect to p_{act}, the introduced redundancy is too high and leads to performance loss, increasing with the gap between the two. Similar considerations hold in case p is underestimated. However, we can appreciate the fact that the performance impairment is limited.

TABLE II
PERFORMANCE SENSITIVITY OF IMAGES LENNA AND GOLDHILL AT RATE 1.2 bpp

Image		Lenna		Goldhill	
p		0.10	0.20	0.10	0.20
p_{act}	0.05	38.3	38.0	34.6	34.3
	0.10	37.9	37.8	34.2	34.0
	0.15	37.6	37.6	33.8	33.8
	0.20	37.3	37.5	33.4	33.6
	0.30	36.9	37.2	33.0	33.3

Future work will include an extension to motion JPEG 2000 and to R-D optimized multiple description schemes for the H.264/AVC video standard.

REFERENCES

[1] E. Baccaglini, T. Tillo, and G. Olmo, "Network adaptive multiple description coding for JPEG2000," in *Proc. IEEE Int. Conf. Image Processing*, Sep. 2005, vol. 3, pp. III-936–III-939.

[2] T. Tillo, E. Baccaglini, and G. Olmo, "A flexible multi-rate allocation scheme for balanced multiple description coding applications," in *Proc. IEEE Int. Workshop Multimedia Signal Processing*, Nov. 2005.

[3] V. K. Goyal, "Multiple description coding: compression meets the network," *IEEE Signal Process. Mag.*, vol. 18, no. 5, pp. 74–93, Sep. 2001.

[4] V. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 821–834, May 1993.

[5] J. Cardinal, "Multistage index assignments for M-description coding," in *Proc. Int. Conf. Image Processing*, Sep. 2003, vol. 3, pp. 249–252.

[6] E. Soyak, Y. Eisenberg, F. Zhai, R. Berry, T. N. Pappas, and A. K. Katsaggelos, "Channel modeling and its effect on the end-to-end distortion in wireless video communications," in *Proc. Int. Conf. Image Processing*, Oct. 2004, vol. 5, pp. 3253–3256.

[7] Y. Wang, M. T. Orchard, V. Vaishampayan, and A. R. Reibman, "Multiple description coding using pairwise correlating transforms," *IEEE Trans. Image Process.*, vol. 10, no. 3, pp. 351–367, Mar. 2001.

[8] V. K. Goyal, J. Kovacevic, R. Arean, and M. Vetterli, "Multiple description transform coding of images," *IEEE Trans. Image Process.*, vol. 9, no. 5, pp. 813–826, May 2000.

[9] W. Jiang and A. Ortega, "Multiple description coding via polyphase transform and quantization selection," in *Proc. SPIE Int. Conf. Visual Communication Image Processing*, Jan. 1999, pp. 998–1008.

[10] T. Tillo and G. Olmo, "A novel multiple description coding scheme compatible with the JPEG 2000 decoder," *IEEE Signal Process. Lett.*, vol. 11, no. 11, pp. 908–911, Nov. 2004.

[11] M. Rabbani and D. Santa-Cruz, "JPEG2000 still-image compression standard," in *Proc. Int. Conf. Image Processing*, Thessaloniki, Greece, Oct. 11, 2001.