

Multiple Description Image Coding with Block-Coherent Redundant Dictionaries ^{*}

Ivana Radulovic and Pascal Frossard

Ecole Polytechnique Fédérale de Lausanne (EPFL)
Signal Processing Institute - LTS4
CH-1015 Lausanne, Switzerland

Abstract. This paper presents a method for multiple description coding, by clustering generic redundant dictionaries into blocks of N atoms. Since these blocks are likely to gather coherent atoms, the encoder can distribute atoms from the same coherent block to different descriptions, without important penalty on the side distortion. In the same time, atoms from the same blocks are sufficiently different, such that they contribute together to efficiently lowering the central distortion. As opposed to state-of-the-art methods that are purely redundant for the most important components, our strategy allows to gain on the central distortion, without loosing on the side distortion. The proposed scheme is applied to the multiple description coding of digital images, and simulation results show increased performances compared to state-of-the-art schemes, in terms of both distortions, and robustness to loss probability variations.

1 Introduction

The goal of multiple description coding (MDC) is the generation of several independently decodable representations of a source signal, in such a way that the distortion of the decoded signal improves with the number of descriptions that are correctly received. While the number of descriptions is a priori not constrained, the most popular scenario in practical applications is represented by two descriptions sent over lossy channels, as shown on Figure 1. Receiving only the description i (with $i = \{1, 2\}$) results in the *side* distortion D_i , while receiving both descriptions induces a *central* distortion denoted D_{12} .

Multiple description coding offers an interesting solution for image transmission over lossy channels, with graceful degradation of the quality when loss rate increases. In the same time, coding schemes based on redundant signal expansions over dictionaries of atoms, provide very interesting alterna-

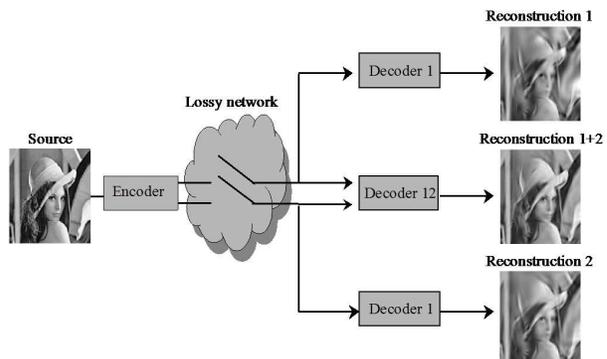


Fig. 1. MDC scenario with two descriptions.

tives for low bit rate image communication applications ([1], [2]). However, the generation of multiple descriptions with redundant transforms has been poorly explored in the context of image coding, although multiple descriptions and redundant dictionaries have inherent common properties that could be efficiently combined.

In [1] the authors propose generation of multiple description video coder based on Matching Pursuit. In their implementation, the first (and most important) atoms are repeated in both descriptions, while the remaining atoms are alternately split between the descriptions. However, the repeated atoms do not bring any improvement in the reconstruction when both descriptions are received. In [3], the authors use *frames* as another redundant expansion method to generate multiple descriptions, and they show that this kind of expansion performs better than Unequal Error Protection (UEP) schemes. However, using frames for the generation of multiple descriptions is still not widely spread due to the fact that not all subsets of received frame components provide a good reconstruction, [4].

Nevertheless, to the best of our knowledge, no attempt has been made on judicious arrangements of atoms, in order to produce multiple balanced descriptions that benefit from the redundancy of the dictionary. In this paper, we propose a method for the generation of N -balanced descriptions, by a

^{*} This work has been supported by the Swiss National Science Foundation grant PP-002-68737. Email: {ivana.radulovic, pascal.frossard}@epfl.ch

priori clustering of the redundant dictionary into blocks of N coherent atoms. The most significant atoms in the signal representation are drawn from the clusters that better approximate the signal, and distributed into the different descriptions. As compared to atom repetition, such a strategy allows for a better central distortion since atoms in different descriptions are not identical; in the same time, it does not penalize the side distortions since atoms from the same cluster are likely to be highly correlated and thus get a similar "importance". A refinement stage finally allocates the least important atoms, alternatively between the descriptions. Experimental results demonstrate the improved performance of our scheme, both in terms of average distortion and the robustness to network changes.

The paper is organized as follows. In Section 2, we explain the method for the generation of multiple descriptions based on dictionary clustering. We describe the generation and reconstruction, as well as the bounds on achievable distortions. Section 3 gives the comparisons of our scheme and state-of-the-art schemes. Finally, Section 4 concludes the paper.

2 Multiple balanced descriptions with redundant dictionaries

2.1 Redundant expansions with Matching Pursuit

The main objective in signal compression is to provide a compact representation with as few elements as possible. Using redundant sets of basis functions may provide an interesting advantage when transforming the signal into its most important features. Redundancy is also expected to be beneficial in the signal reconstruction, when information is lost.

Searching for the sparsest signal representation in an overcomplete dictionary of functions, is in general a NP-hard problem, [5]. However, it is usually sufficient to find a nearly optimal solution that would reduce the search complexity in a great manner. Matching Pursuit (MP), introduced by Mallat in [6], is an example of a greedy algorithm that iteratively decomposes any function s in the Hilbert space \mathcal{H} with the elements from a redundant dictionary called *atoms*. Let all the atoms, denoted by a_i , have a unit norm, $\|a_i\|_2 = 1$, and let $\mathcal{D} = \{a_i\}$, $i = 1, 2, \dots, |\mathcal{D}|$. By setting $R_0 = s$, the signal is first decomposed as:

$$R_0 = \langle a_0, R_0 \rangle a_0 + R_1 \quad (1)$$

where a_0 is chosen so as to maximize the correlation with R_0 :

$$a_0 = \arg \max_{\mathcal{D}} |\langle a_i, R_0 \rangle| \quad (2)$$

and R_1 is the residual signal after the first iteration. The algorithm is iteratively carried out, by applying the same procedure to the residual signal. After M iterations, the signal s is decomposed as follows:

$$s = \sum_{i=0}^{M-1} \langle R_i, a_i \rangle a_i + R_M \quad (3)$$

2.2 Dictionary clustering

Matching Pursuit generally results in very sparse signal expansions, where the first elements capture most of the signal energy. Multiple description coding should ideally provide redundancy for these important signal components. On one side, the descriptions should however be different enough so that the reconstruction quality improves with the number of description. On the other side, each description needs to provide an acceptable reconstruction quality when received alone.

To reach this objective, we propose an a priori stage that clusters a redundant dictionary, such that each of the clusters contains the N correlated atoms. Now, instead of running a pursuit on the level of atoms, we could iteratively decompose the signal on the level of clusters. Once the best cluster is chosen, we can attribute its atoms to different descriptions. Receiving more descriptions allows for a combination of correlated atoms in order to get the one that better approximates a signal.

In general, we can cluster a dictionary using either an *top-down* or a *bottom-up* approach. While the former method fixes a priori the number of clusters, [7], the latter one allows us to set the cardinality of each cluster, [8]. Since each of our clusters should have N atoms for N descriptions, we will use the *bottom-up* approach.

Once the clusters are made, each of them can be subsequently represented as a molecule, taken as the centroid of its atoms. The centroid m_i minimizes the sum of the distances to all the atoms in the cluster i :

$$m_i = \arg \min_{m, \|m\|=1} \sum_j d(m, a_j) \quad (4)$$

Different metrics can be used for the distance measure; one of the most popular ones is : $d(a_i, a_j) = 1 - |\langle a_i, a_j \rangle|^2$. If the atoms are strongly correlated, their distance is close to 0, while in the case of orthogonal atoms this distance is 1.

2.3 Generation of N descriptions

While providing multiple correlated atoms for each step of the iterative overcomplete expansion is crucial for the most energetic components of the signal, it becomes less attractive in the refinement steps, where energy of atoms becomes small. In this case, coding with a lot of redundancy results in a waste of bandwidth resources. The most efficient joint source and channel coding schemes proceed by unequal error protection, and we pursue the same idea here.

We propose a two-stage coding for obtaining multiple descriptions, where the most important features are first redundantly coded between descriptions. Instead of repeating the most important atoms in the first stage of coding, we however distribute very similar, but not identical, atoms in different descriptions. The low energy atoms are then simply distributed alternatively between descriptions.

In the first stage of coding, we propose the representation of the signal to be computed on the molecules, instead of the atoms themselves. The L molecules m_i that best approximate the signal s are selected by running Matching Pursuit on the set of molecules, which yields:

$$s = \sum_{j=0}^{L-1} \langle R_j, m_j \rangle m_j + R_L \quad (5)$$

The multiple descriptions are built by distributing each atom from the blocks corresponding to these molecules, into different descriptions. Formally, if a molecule m_j is chosen in the j -th stage of MP, we attribute its child a_i^j to description i , with $i = 1, 2, \dots, N$.

In the second stage, after the L most significant molecules have been chosen, we run MP on the residual signal R_L , at the level of atoms. The chosen atoms are alternately distributed between the descriptions, to eventually generate descriptions with a total of M atoms. The atoms in description i can then be represented by a generating matrix $\Phi_i = \{a_i^j\}, j = 1, 2, \dots, M$, where the first L rows of Φ_i are children of the L selected molecules, and the remaining $M-L$ rows correspond to atoms that are alternately distributed between descriptions.

To generate description i , the signal is finally projected onto Φ_i :

$$\Phi_i s^T = C_i \quad (6)$$

where C_i gives the contribution of each atom in Φ_i . When all atom contributions have been computed,

C_i is uniformly quantized into \tilde{C}_i . Together with indexes of atoms in Φ_i , \tilde{C}_i are then entropy coded to form the i -th description.

2.4 Reconstruction

The encoding scheme proposed in the previous section does not conserve the energy of the signal, which cannot be reconstructed by simple linear combination of the atoms as they are received at the decoder. We therefore need to design a decoding process that removes the redundancy that has been introduced in the encoding stage.

We consider that descriptions are balanced (which is generally the case by the proposed construction method) so that any set of k descriptions induces a similar distortion $D_{12\dots k}$, $1 \leq k \leq N$. If only a description i is received, we reconstruct the signal simply by inverting the operations done at the encoder, as follows:

$$r_i = (\Phi_i^\dagger \cdot \tilde{C}_i)^T \quad (7)$$

where T and \dagger respectively denote the transpose and pseudoinverse matrices. Such a reconstruction induces a distortion D_i that can be bounded by:

$$D_i = \frac{\|s - r_i\|^2}{S} \leq \frac{\|s\|^2}{S} - \frac{\sum_{j=1}^M \tilde{C}_i^2(j)}{S(1 + \alpha(M-1))} \quad (8)$$

where S corresponds to the signal size, and α is the maximal inner product between any two atoms in Φ_i . If k descriptions are received, the signal is reconstructed in a similar way:

$$r_{12\dots k} = [[(\Phi_1 \dots \Phi_k)^T]^\dagger \cdot [\tilde{C}_1 \dots \tilde{C}_k]]^T \quad (9)$$

Since the matrix $\Phi = [\Phi_1 \dots \Phi_k]$ has dimensions $2M \times M$, computing its pseudoinverse is quite involved. However, the complexity can be drastically reduced using the fact that $\Phi^\dagger = \Phi^T (\Phi \cdot \Phi^T)^\dagger$. Namely, instead of computing a pseudoinverse of Φ , we simply compute the inverse of $\Phi \cdot \Phi^T$ that is a symmetric $M \times M$ matrix. The distortion $D_{12\dots k}$ after reconstruction, in this case, can be bounded by:

$$D_{12\dots k} = \frac{\|s - r_{12\dots k}\|^2}{S} \leq \frac{\|s\|^2}{S} - \frac{k \cdot \|s\|^2 - k D_1}{S(1 + (k-1)\alpha)}$$

It is clear from the previous expression that the distortion decreases with the number of received descriptions. However, it is also interesting

to notice that for $k \geq 1$: $D_{12\dots k+1} - D_{12\dots k} < D_{12\dots k} - D_{12\dots k-1}$, i.e., the decrease in distortion gets smaller as more descriptions are received. Besides, lower α causes a higher decrease in distortion.

Finally, if we assume that the probability of losing each of the descriptions is equal to p , we can write the average distortion in the presence of losses as:

$$D_{av} = \sum_{k=0}^N \binom{N}{k} p^k (1-p)^{N-k} D_{12\dots k} \quad (10)$$

3 Experimental Results

For the generation of multiple image descriptions, we used an overcomplete dictionary of edge-like functions, as proposed in [2]. We additionally use a bottom-up Tree Based Pursuit algorithm [8], which implements a clustering strategy where a fixed number N of similar atoms are grouped together. We compare the performance of our scheme with the existing MDC schemes such as atom sharing scheme, [1], and the FEC scheme based on UEP, [9], depicted on Figure 2. The atom sharing scheme repeats a certain number of most important atoms in all the descriptions, while the remaining atoms are alternatively split between descriptions. On the other side, FEC scheme applies a systematic code, like a Reed-Solomon code, column-wise across the N -packet block. Here, different sets of atoms are protected according to their importance.

First, we compare the performance of our scheme for $N = 2$ descriptions with the existing schemes. Note that, when $N = 2$, the UEP scheme is equal to atom sharing scheme. For the comparison, we first examine how the minimal achievable average distortion depends on the packet loss ratio in the network, p . To do so, we find the optimal number of shared atoms for the atom sharing scheme and the optimal L for our scheme. We attribute 200 atoms to each packet/description and approximately 25 bits/atom (11 bits coding the atom index plus 14 bits for the uniform quantization of coefficients). Therefore, the total rate becomes 1.25 kB. The results are shown on Figure 3. We can see that our scheme provides improvement of up to 1.6 dB comparing to the atom sharing (and UEP) scheme. This is due to the fact that our scheme takes advantage from all the received atoms, while the existing schemes cannot use the redundant atoms.

Next, we compare all schemes optimized for a given loss ratio p , but when conditions in the network change. Figure 4 shows the performance of

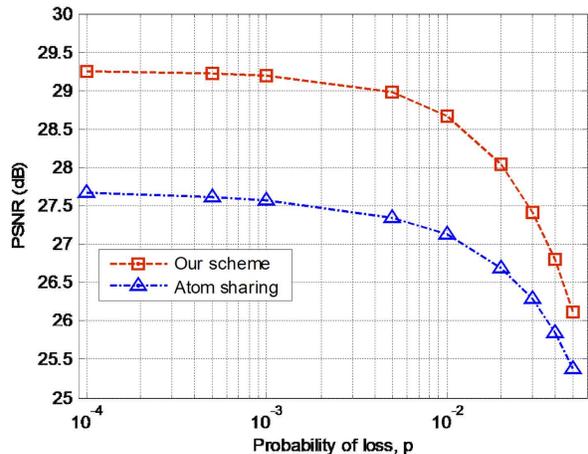


Fig. 3. PSNR vs loss probability for our and atom sharing scheme, optimized for a total rate of 1.25 kB.

both schemes optimized for $p = 10^{-3}$ in the whole range of losses. We can see that our scheme gives up to 1.3 dB higher PSNR at low loss ratios, while at very high losses it still performs better, but very similar to existing schemes.

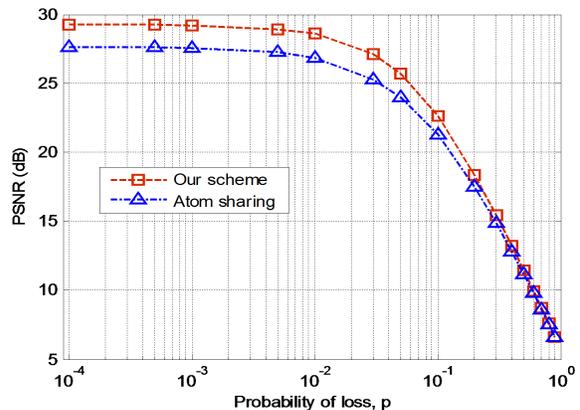


Fig. 4. PSNR vs packet loss ratio, for our and atom sharing scheme, optimized for a total rate of 1.25 kB, and a loss probability of 10^{-3} .

Further on, we examine how the reconstruction looks as a function of number of received descriptions. All schemes are again optimized for $p = 10^{-3}$ and the results are depicted on Figures 5 (a) and (b). We can observe that the side reconstruction in our scheme has 1.7 dB higher PSNR, while the central reconstruction is slightly better (0.5 dB).

Now, we consider the case of $N = 3$ descriptions and do similar comparisons like in the case of two descriptions. The minimal average distortion as a function of p for the three schemes is given in Figure 6. We see that our scheme outperforms the existing scheme in the whole range of losses, especially

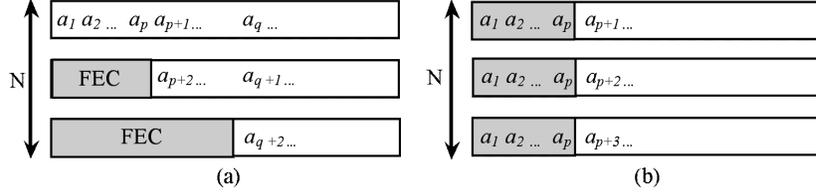


Fig. 2. (a) FEC scheme and (b) atom sharing scheme



Fig. 5. Reconstructed Lena, as a function of a number of received descriptions: (a) our scheme, (b) FEC scheme

Fig. 7. Reconstructed Lena, as a function of a number of received descriptions: (a) our scheme, (b) FEC scheme

at low packet loss ratios, where the improvement is up to 1.7 dB. It is also interesting to notice that the FEC and atom sharing scheme perform similarly at low losses, while there is up to 0.7 dB higher PSNR for the FEC scheme when $p = 5\%$. This is due to the possibility of FEC scheme to protect strongly the most important atoms. Figures 7 (a)

number of received descriptions. We can observe that receiving only one description results in 1.8 dB higher PSNR with our scheme, while the FEC scheme gives 1.5 higher PSNR when two descriptions are received. Finally, when all descriptions are received, our scheme will give slightly better PSNR than a FEC one (0.1 dB).

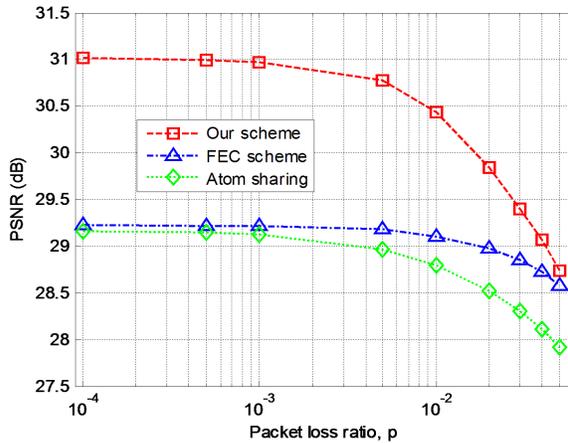


Fig. 6. PSNR vs packet loss ratio, for our, FEC and atom sharing scheme, optimized for a total rate of 1.87 kB, and a loss probability of 10^{-3} .

and (b) show the reconstructions for our and FEC scheme optimized for $p = 1\%$, as a function of the

To explore further the performance of our scheme, we fix the total number of atoms to 600 and vary the number of descriptions from 2 to 4, which means that we vary the number of atoms per description from 300 to 150. Minimal average distortions as a function of p for $N = 2, 3$ and 4 are shown in Figure 8. We can see that in the range of very low loss ratios (up to 10^{-3}) it is optimal to have two descriptions. As the losses increase from $p = 10^{-3}$ to $p = 10^{-2}$, having three descriptions will give the smallest distortion. If the losses exceed 1%, the optimal would be to have four descriptions, in which case the reconstruction gives 0.5 dB higher PSNR. The fact that the optimal number of descriptions increases as the losses increase is not a surprising result and it has been already reported in [10] and [11], where the multiple description scalar and vector quantization were used to produce an arbitrary number of descriptions. Therefore, we have again shown that having only two descriptions is not al-

ways optimal. Finally, we compare our optimized

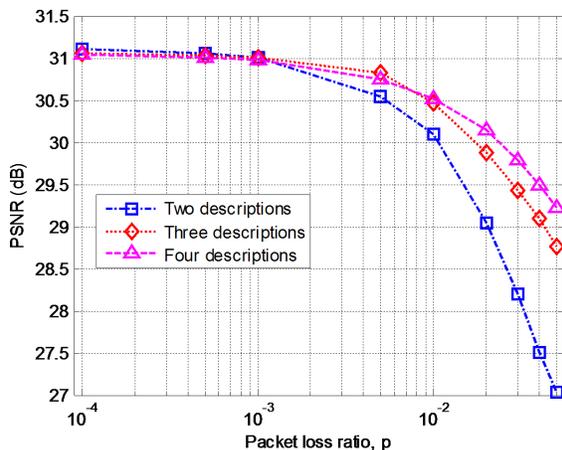


Fig. 8. Distortions, as a function of packet loss ratio and number of descriptions.

scheme (obtained by choosing the optimal number of descriptions for each p) with the optimized FEC scheme, but which works with the coefficients obtained by projection (6), see Figure 9. We can see that our scheme still outperforms even the new FEC scheme up to 0.6 dB and in the range of losses up to $p = 1\%$. The reason why FEC performs worse at low losses is the higher amount of protection it put to the atoms; this is also why it performs better than our scheme at high losses, since the most important atoms are heavily protected.

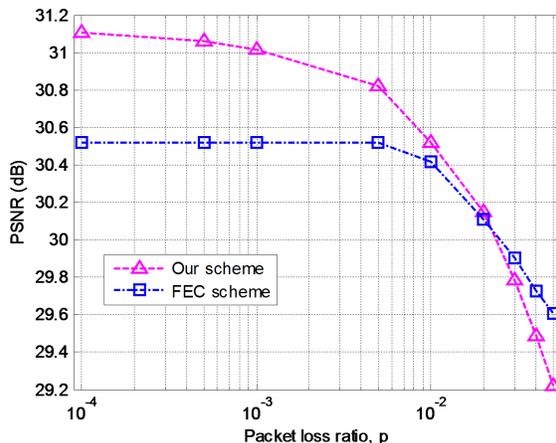


Fig. 9. Comparison of our optimized scheme with the optimized FEC scheme.

4 Conclusions

In this paper we have shown how the redundant dictionaries can be exploited to generate multiple descriptions of images. In contrast to state-of-the-art schemes, we do not repeat the most important

atoms, but instead send very similar but not identical atoms in each description. We showed that our method gives better performance both in terms of the side and central distortion, and that is more robust to network changes.

References

1. X. Tang and A. Zakhor, "Matching pursuits multiple description coding for wireless video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, June 2002, pp. 566–575.
2. R. Figueras i Ventura, P. Vandergheynst and P. Frossard, "Low rate and flexible image coding with redundant representations," *IEEE Transactions on Image Processing*, March 2006.
3. V.K. Goyal, J. Kovacevic, R. Arean, M. Vetterli,, "Multiple description transform coding of images," in *Proceedings of the IEEE International Conference on Image Processing, Oct. 1998*, pp.674 - 678.
4. V.K. Goyal, J. Kovacevic, J.A. Kelner, "Quantized frame expansions with erasures," *Applied and Computational Harmonic Analysis*, May 2001, Vol. 10, pp. 203-233.
5. G. Davis, S. Mallat and M. Avellaneda, "adaptive greedy approximations", *Journal of Constructive Approximation*, vol. 13, pp. 57-98, 1997.
6. S. Mallat and Z. Zhang, "Matching pursuit with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.
7. I.S. Dhillon, E.M. Marcotte and U. Roshan, "Di-metrical clustering for identifying anti-correlated gene clusters," *Bioinformatics*, vol. 19, no. 13, pp. 1612–1619, 2003.
8. P. Jost, P. Vandergheynst and P. Frossard, "Tree-based pursuit: Algorithm and properties," *accepted for publication in IEEE Transactions on Signal Processing*.
9. P. Frossard and P. Vandergheynst, "Unequal error protection of atomic image streams [submitted paper]," *IEEE Transactions on Image Processing*, 2005.
10. I. Radulovic and P. Frossard, "Fast index assignment for balanced n-description scalar quantization," in *Proceedings of the IEEE Data Compression Conference, Mar. 2005*, pp. 474.
11. J. Ostergaard, J. Jensen, R. Heusdens, "n-Channel Symmetric Multiple Description Lattice Vector Quantization," in *Proceedings of the Data Compression Conference, Mar. 2005*, pp.378 - 387.