Multiple description image coding with redundant expansions and optimal quantization

Ivana Radulovic and Pascal Frossard

Ecole Polytechnique Fédérale de Lausanne (EPFL)

Signal Processing Institute - LTS4

CH-1015 Lausanne, Switzerland

Abstract— This paper addresses the problem of optimal rate allocation for multiple description coding with redundant signal expansions. In case of redundant descriptions, the quantization of the transform coefficients has clearly to be adapted to the importance of the basis functions, to the redundancy in the representation, and to the expected loss probability on the transmission channel. We derive a rate-distortion optimal solution for the scalar quantization of coefficients in redundant signal representations. The application of the optimal rate allocation to a typical image communication problem demonstrates performance gains with respect to scheme based on uniform quantization with fixed step size, and to solutions based on unequal error protection.

I. INTRODUCTION

Multiple Description Coding (MDC) is an efficient solution for error-resilient transmission of multimedia streams on lossy packet networks. It also offers an interesting alternative in distributed delivery schemes, where the multimedia application takes benefit of the diversity offered by the network infrastructure. Basically, the idea behind MDC is to build several independent representations (e.g., data packets) of a source in such a way that each of them provides an acceptable quality for the signal reconstruction, while this quality improves as more descriptions become available at decoder.

The design of efficient image and video multiple description coding schemes is not a trivial problem, as they have to offer a well-chosen trade-off between compression efficiency and error resiliency. Redundant signal transforms represent one of the most promising alternatives for multiple description coding. They provide good energy compaction properties, resilience to additive noise and quantization and greater freedom to capture important signal characteristics [1]. The redundancy that remains after signal transformation can mitigate the effect of errors or erasures, since it offers the possibility to estimate the lost elements based on the received ones.

Although MDC and redundant transforms have many inherent properties in common, little work has been done to combine them. Previous works propose the application of redundant wavelet transforms [2], frames or Matching Pursuit expansions for multiple description image coding. In particular, frame expansions have been proposed in [3], [4], where the authors demonstrate better performance with MDC based on harmonic frames, in comparison to schemes based on unequal error protection. However, the use of frames for MDC is limited by the fact that not all subsets of frame components are able to offer a good signal reconstruction, [5]. Matching Pursuit has been proposed in [6] for the generation of two descriptions of video sequences. The elements of a redundant dictionary (i.e., atoms) that best approximate the signal are repeated in both descriptions, while the remaining atoms of the decomposition are alternatively split between the descriptions. The redundancy between the descriptions is controlled with a number of shared atoms. However, if no loss occurs, such a repetition results in an obvious waste of resources.

In order to avoid the replication of signal information in different descriptions, we proposed in [7] to construct an arbitrary number of descriptions by partitioning the dictionary of atoms into clusters of similar elements. When the signal is well approximated by the elements of one cluster, the multiple description coding scheme distributes these atoms in different descriptions. It ensures that descriptions are correlated and, in the same time, all atoms contribute in improving the approximation of the signal, since they are still different. The signal reconstruction has to integrate several redundant components, and the signal reconstruction can be posed as an optimization problem whose constraints are given by the set of atoms and their relative contribution, which are available at the decoder. Due to redundancy between received components, uniform scalar quantization of the atom coefficients [5], [7], [8] is clearly suboptimal in a rate-distortion sense. The definition of optimal quantization of coefficients in the particular case of Gaussian sources has been proposed in [9].

In this paper, we design a quantization scheme for the coefficients of redundant atoms in multiple description image coding. We concentrate on scalar quantization in order to avoid the complexity imposed by vector quantization. We propose to vary the quantization step sizes for the different coefficients, depending on the redundancy in the descriptions, and the expected probability of erasures. We design an optimal uniform scalar quantizer, which tends to finely quantize the coefficients of atoms that bring most of the signal energy. In the same time, when all the descriptions are likely to be available at the decoder, the redundancy between descriptions decreases, and the quantization step sizes are adapted accordingly. We apply the optimal scalar quantization to multiple description image coding for typical transmission scenarios. As we will show, this scheme offers improved performance when compared to uniform quantization with fixed step size, or to FEC-based MDC strategies.

II. Multiple description coding with redundant expansions

The multiple description coding problem that we consider in this paper consists of generating N descriptions of a signal s of size S. The signal lies in a real S-dimensional vector space, $\mathcal{R}^S$, endowed with a real-valued inner product. The signal s is to be represented with a finite collection of unitary norm elementary signals or basis functions, called the *atoms*. Denote by $\mathcal{D} = \{a_i\}_{i=1}^{|\mathcal{D}|}$ such a collection of $|\mathcal{D}|$ atoms, that we call a *dictionary*. Redundant dictionaries in general are such that $|\mathcal{D}| \gg d$. We do not set any particular constraint on the dictionary design, except that it should span the entire signal space.

Signal descriptions are formed by selecting atoms from $\mathcal{D}$, along with their respective contribution in the signal reconstruction. In other words, we form a description i by selecting M atoms from $\mathcal{D}$, which can be gathered in the generating matrix $\Phi_i = \{a_{ij}\}$, with $j = 1, 2, ..., M$. The atom coefficients in description i are simply given by the projection of the signal s onto the generating matrix Φ_i, as:

$$\Phi_i s^T = C_i \ , \qquad (1)$$

where C_i is a vector of real values $c_{ij} = \langle s, a_{ij} \rangle$, with $j = 1, 2, ..., M$. We assume that each coefficient c_{ij} undergoes a uniform scalar quantization with a stepsize Δ_{ij}, before coding and transmission along with the indexes of the atoms in Φ_i. The quantized coefficient $\hat{c}_{ij}$ can be written as $\hat{c}_{ij} = c_{ij} + \eta_{ij}$, where η_{ij} denotes the quantization noise. Similarly, we denote by $\hat{C}_i$ the vector of quantized coefficients.

The descriptions are sent over a lossy channel, and the decoder eventually reconstructs the signal with the descriptions that have not been erased. If only the description i is available at the decoder, the signal is reconstructed as

$$r_i = (\Phi_i^\dagger \cdot \hat{C}_i)^T,$$

where $\dagger$ denotes the pseudo-inverse matrix. We now define the matrix Ω_i as $\Omega_i = (\Phi_i \Phi_i^T)^\dagger \cdot \Phi_i = \Phi_i \cdot (\Phi_i^T \Phi_i)^\dagger$. The jth row of the matrix can be written as $\omega_{ij} = a_{ij} \cdot (\Phi_i^T \Phi_i)^\dagger$. The reconstructed signal can then be rewritten as

$$r_i = \sum_{j=1}^{M} \hat{c}_{ij} \omega_{ij} = \sum_{j=1}^{M} (c_{ij} + \eta_{ij}) \omega_{ij}. \qquad (2)$$

It follows that the MSE distortion resulting from receiving only the description i can be expressed as:

$$
\begin{aligned}
D_i &= \frac{\|s - \sum_{j=1}^{M}(c_{ij} + \eta_{ij})\omega_{ij}\|^2}{S} \\
&\leq \frac{\sum_{j=1}^{M}\|s - c_{ij}\omega_{ij}\|^2}{S} + \frac{\sum_{j=1}^{M}\Delta_{ij}^2\|\omega_{ij}\|^2}{12S}. \quad (3)
\end{aligned}
$$

The first term of the distortion upper-bound corresponds to the source distortion. The source distortion can be non-negligible when the coding rate is kept low, or when the number of atoms per description stays small. The second term of the upper-bound corresponds to the distortion due to quantization, that is influenced by term $\|\omega_{ij}\|^2$, which reflects the properties of Φ_i, and the redundancy between the atoms that form the description i.

Similarly, if the decoder receives a set K of $k \geq 2$ descriptions of M atoms, the reconstructed signal can be written as :

$$r_K = (\Phi_K^\dagger \cdot \hat{C}_K)^T = \sum_{j=1}^{kM}(c_{Kj} + \eta_{Kj})\omega_{Kj}, \qquad (4)$$

where Φ_K is the matrix constructed by gathering the generating matrices of the k descriptions available at the decoder, and $\hat{C}_K$ is the corresponding vector of quantized coefficients. If we define a matrix Ω_K similarly to Ω_i, the distortion can be expressed as:

$$
\begin{aligned}
D_K &= \frac{\|s - \sum_{j=1}^{kM}(c_{Kj} + \eta_{Kj})\omega_{Kj}\|^2}{S} \\
&\leq \frac{\sum_{j=1}^{kM}\|s - c_{Kj}\omega_{Kj}\|^2}{S} + \frac{\sum_{j=1}^{kM}\Delta_{Kj}^2\|\omega_{Kj}\|^2}{12S}. \quad (5)
\end{aligned}
$$

III. Optimal coefficient quantization

In a multiple description coding framework, the optimal coefficient quantization consists of minimizing the *average* quantization error in the presence of losses, for a given bit budget allocated to coding the coefficients. The average distortion due to quantization is given as the sum of distortions resulting from all the possible realizations at the decoder, weighted by their respective probabilities:

$$\bar{D}_q = \sum_{K=1}^{2^N-1} p_K\, D_K^q = \frac{1}{12S} \sum_{K=0}^{2^N-1} p_K \sum_{j=1}^{kM} \Delta_{Kj}^2\|\omega_{Kj}\|^2, \quad (6)$$

where D_K^q represents the distortion due to quantization when the set of descriptions K is available at the decoder (i.e., the second term of the above distortion upper-bounds). In the same time, if the coefficient c_{ij} is uniformly quantized in the interval $[-\|s\|, \|s\|]$ with a step size Δ_{ij}, the minimal rate required for coding all the coefficients can be written as:

$$R_q = \sum_{i=1}^{N} \sum_{j=1}^{M} log_2 \frac{2\|s\|}{\Delta_{ij}}. \qquad (7)$$

The objective of optimal quantization is to define the quantization step sizes Δ_{ij} such that $\bar{D}_q$ is minimized for a given R_q. It is equivalent to minimizing the Lagrangian function $\mathcal{L}(\lambda) = \bar{D}_q + \lambda R_q$. The optimal quantization step-sizes can be derived by differentiating $\mathcal{L}$ with respect to Δ_{ij}, and setting the partial derivative $\frac{\partial \mathcal{L}(\lambda)}{\partial \Delta_{ij}}$ to zero.

In the case where the encoder generates two descriptions ($N = 2$), which can be lost independently with a probability

p, the average quantization distortion can be written as:

$$\bar{D}_q = (1-p)^2 D_K^q + p(1-p)(D_1^q + D_2^q)$$

$$= \frac{(1-p)^2}{12S}\left(\sum_{j=1}^{M}\Delta_{1j}^2\|\omega_{Kj}\|^2 + \sum_{j=1}^{M}\Delta_{2j}^2\|\omega_{Kl}\|^2\right)$$

$$+ \frac{p(1-p)}{12S}\left(\sum_{j=1}^{M}\Delta_{1j}^2\|\omega_{1j}\|^2 + \sum_{j=1}^{M}\Delta_{2j}^2\|\omega_{2j}\|^2\right) \quad (8)$$

where K represents the set formed with both descriptions and $l = j + M$. By solving the above optimization problem, we finally obtain the following expressions for the quantization step sizes:

$$\Delta_{1j}^2 = \frac{2^A}{(1-p)^2\|\omega_{Kj}\|^2 + p(1-p)\|\omega_{1j}\|^2} \quad (9)$$

$$\Delta_{2j}^2 = \frac{2^A}{(1-p)^2\|\omega_{Kl}\|^2 + p(1-p)\|\omega_{2j}\|^2}, \quad (10)$$

where

$$A = \frac{2Mlog_2 2\|s\| + Mlog_2(1-p) + \frac{1}{2}log_2\pi_\omega - R_q}{M}$$

and

$$\pi_\omega = \prod_{j=1}^{M}((1-p)\|\omega_{Kj}\|^2 + p\|\omega_{1j}\|^2)\prod_{j=1}^{M}((1-p)\|\omega_{Kl}\|^2 + p\|\omega_{2j}\|^2).$$

This example highlights that the optimal step sizes are dependent on the atoms, on the redundancy among atoms, and on the loss probability p. Note that all the coefficients equally contribute to the distortion, as it can be expected from optimal rate allocation.

IV. Multiple description image coding

This section now describes in more details how multiple descriptions of an image can be generated with redundant dictionaries of functions. We initially construct an overcomplete dictionary with both edge-like and isotropic functions, in order to offer an efficient representation of a large class of natural images [10]. The first function, an isotropic Gaussian $2-D$ function, is responsible for efficient representation of the low frequency characteristics of an image. The second one, an anisotropic function is used to capture the contours that represent most of the content of natural images. It is built on a Gaussian along one direction and a second derivative of a Gaussian along the orthogonal direction. Geometric transforms (translation, rotation and scaling) are then applied to both functions to build a structured redundant dictionary. We allow the translation parameters to be any integers smaller than the image size. For the first function, the scaling is isotropic and varies from $\frac{1}{32}$ to $\frac{1}{4}$ of the image size on a logarithmic scale with a resolution of one third of octave. For the second one, the scaling is anisotropic, and the scale parameters are uniformly distributed on a logarithmic scale from one to $\frac{1}{8}$ of the image size, with a resolution of one third of octave. We also allow the rotation parameter to vary in increments of $\frac{\pi}{18}$.

Similarly to the algorithm proposed in [7], the N descriptions are generated by partitioning the redundant dictionary of atoms into clusters of N similar atoms, each of them being further represented by a particular function that we call a *molecule*. A molecule reflects the characteristics of the atoms within a cluster, and can be computed for example, as a weighted sum of the N atoms of the cluster. The signal decomposition is then performed with molecules, and a Matching Pursuit search iteratively determines the molecules that best approximate the image to be encoded. Once a molecule is selected, its N children atoms are distributed among the N descriptions. When the L most significant molecules have been identified, a residual signal is built by subtracting the reconstruction based on all the selected molecules, from the original image. A Matching Pursuit expansion of the residual signal is eventually performed directly on atoms. The atoms are simply distributed alternatively between descriptions, to eventually generate descriptions with a total of M atoms. Upon completing both stages, the M atoms in description i are gathered in a generating matrix $\Phi_i = \{a_{ij}\}$, with $j = 1, 2, ..., M$, 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 alternatively distributed between descriptions. In addition to atoms, the descriptions also carry coefficients that reflect the relative contribution of each atom in the signal reconstruction. The coefficients are computed by projection of the signal s onto the matrix Φ_i. Finally, the coefficients are quantized with the optimal rate allocation strategy proposed in the previous section.
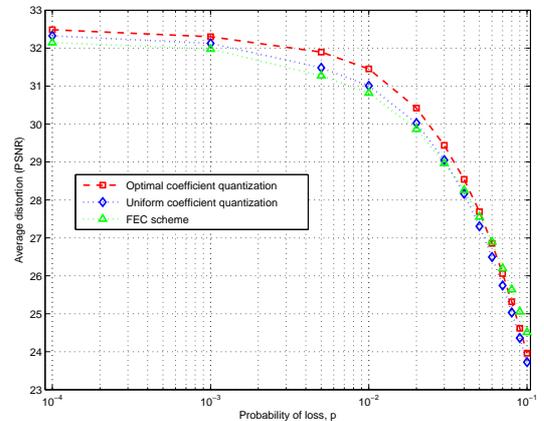
V. Performance evaluation

Fig. 1. Minimal achievable average distortions, as a function of p, for two descriptions of the Lena image (128x128).

We now compare the performance of the proposed solution to the performance of a scheme using uniform quantization with a fixed step size, and to an MDC scheme based on unequal error protection of atoms [11]. We first fix the total number of atoms per descriptions to 400, so that the size of each description is smaller than packets in typical multimedia communication scenarios. The fixed stepsize for the uniform
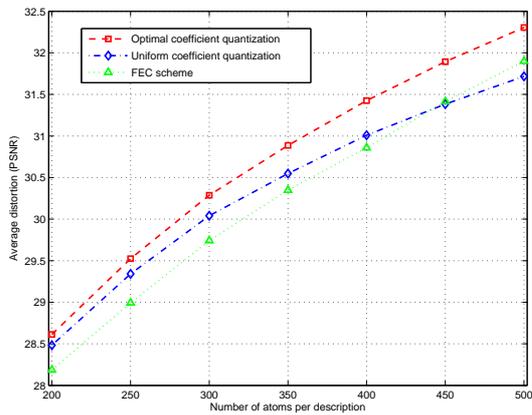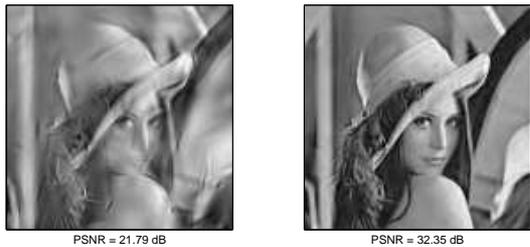
Fig. 2. Minimal achievable average distortions, as a function of p, for two descriptions of the Lena image (128x128).

PSNR = 20.95 dB   PSNR = 31.93 dB

(a) Uniform quantization with fixed stepsize.

PSNR = 21.79 dB   PSNR = 32.35 dB

(b) Optimal quantization.

Fig. 3. Reconstructed Lena (128 x 128). Side reconstruction from the first description (left column) and central reconstruction (right column).

quantization scheme is set to $\Delta = 10$, which leads to the total rate of 8.8 kbits per description (11 bits for coding the atom indices plus 11 bits for the quantization). Figure 2 shows the minimal achievable average distortion when the packet loss ratio on the network varies from 10^{-4} to 10^{-1}. We can see that using an optimal scalar quantization of coefficients can give an improvement of up to 0.45 dB compared to the case of the uniform quantization, and up to 0.6 dB compared to the FEC scheme for low loss probabilities. This is due to the fact that the optimized quantization scheme tends to finely quantize the most important atoms, while it allocates bigger step sizes to coefficients from the less important atoms. Figure 3 shows the side and central reconstructions for the uniform and optimal quantization schemes, when the number of received descriptions is one and two, respectively. In this specific experiment, both schemes are optimized for $p =$
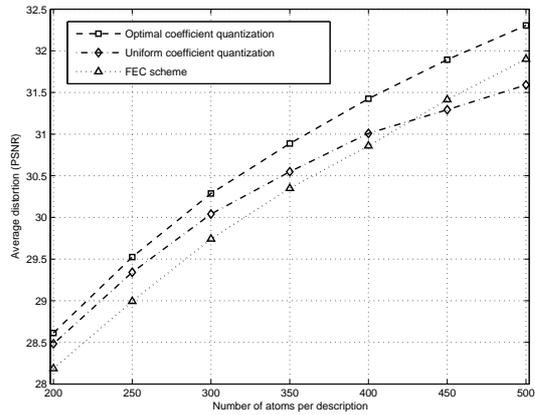

Fig. 4. Minimal achievable average distortions, as a function of number of atoms per description.

1%. We can see that receiving one description brings an improvement of 0.8 dB for the optimal quantization scheme, while the improvement in the central distortion is 0.45 dB.

Finally, Figure 4 shows the performance of the three approaches when the number of atoms per description varies from 200 to 500, and when p is set to 1%. We can see that the optimal quantization scheme always offers the best performance (up to 0.6 dB and 0.57 dB compared to the uniform quantization scheme and the FEC scheme). The gain with respect to uniform quantization with fixed stepsize (11 bits/coefficients) increases when the number of atoms per description increases, as fixed quantization leads to a waste of bits on the lowest energy atoms.

VI. CONCLUSIONS

We have discussed the optimal rate allocation for the quantization of coefficients obtained by multiple description coding with redundant signal transforms. The optimal quantization strategy takes into account the importance of atoms in the signal representation, the redundancy between atoms, and the expected loss probability on the network. The proposed scheme has been applied to a typical image communication scenario, where it improves on uniform quantization with fixed step size, and on unequal error protection of atoms. It clearly highlights the importance of a careful design of the quantization scheme when a signal is coded into multiple redundant descriptions.

REFERENCES

[1] Vivek K. Goyal and Jelena Kovacevic, "Quantized frame expansions with erasures," in *Journal of Computational Harmonic Analysis, Oct. 2001, pp. 203-233*.
[2] C. Tillier, T. Petrisor, B. Pesquet-Popescu, "A motion-compensated overcomplete temporal decomposition for multiple description scalable video coding," *EURASIP Journal on Image and Video Processing*, 2007.
[3] V.K. Goyal, J. Kovacevic, R. Arean and M. Vetterli,, "Multiple description transform coding of images," in *Proceedings of IEEE International Conference on Image Processing*, Oct. 1998, pp. 674–678.
[4] S.S. Channappayya, L. Joonsoo; R.W. Heath Jr. and A.C. Bovik, "Frame based multiple description image coding in the wavelet domain," in *Proceedings of IEEE International Conference on Image Processing*, vol. 3, Sep. 2005, pp. 920–923.

[5] V.K. Goyal, J. Kovacevic and J.A. Kelner, "Quantized frame expansions with erasures," *Journal of Applied and Computational Harmonic Analysis*, vol. 10, no. 3, pp. 203–233, May 2001.

[6] X. Tang and A. Zakhor, "Matching pursuits multiple description coding for wireless video," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 12, no. 6, pp. 566–575, June 2002.

[7] I. Radulovic and P. Frossard, "Multiple description coding with redundant expansions and application to image communications," *EURASIP Journal on Image and Video Processing*, vol. 2007, no. ID 24863, 2007.

[8] G. Rath, C. Guillemot, "Frame-Theoretic Analysis of DFT Codes with Erasures ," *IEEE Trans. on Signal Processing*, vol. 52, no. 2, pp. 447 – 460, Feb. 2004.

[9] S. Mehrotra and P.A. Chou, "On optimal frame expansions for multiple description quantization," in *Proceedings of IEEE International Symposium on Information Theory*, June 2000, p. 176.

[10] R. Figueras i Ventura, P. Vandergheynst and P. Frossard, "Low rate and flexible image coding with redundant representations," *IEEE Trans. on Image Processing*, vol. 15, no. 3, Mar. 2006.

[11] P. Frossard, "On the joint source and channel coding of atomic image streams," in *Proceedings of PCS*, December 2004.