A Posteriori Quantized Matching Pursuit

Pascal Frossard and Pierre Vandergheynst
Signal Processing Laboratory
Swiss Federal Institute of Technology
CH - 1015 Lausanne, Switzerland

Abstract

This paper studies quantization error in the context of Matching Pursuit coded streams and proposes a new coefficient quantization scheme taking benefit of the Matching Pursuit properties. The coefficients energy in Matching Pursuit indeed decreases with the iteration number, and the decay rate can be upper-bounded with an exponential curve driven by the redundancy of the dictionary. The redundancy factor is therefore used to design an optimal *a posteriori* quantization scheme for multi-resolution Matching Pursuit coding. Bits are optimally distributed between successive coefficients according to their relative contribution to the signal representation. The quantization range and the number of quantization steps are therefore reduced along the iteration number. Moreover, the quantization scheme selects the optimal number of Matching Pursuit iterations to be coded to satisfy rate constraints. Finally, the new exponentially upper-bounded quantization of Matching Pursuit coefficients clearly outperforms classical uniform quantization methods for both random dictionaries and Gabor dictionaries in the practical case of image coding.

I. Introduction

Non-orthogonal transforms presents several interesting properties which position them as an interesting alternative to orthogonal transforms like DCT or wavelet based schemes. Decomposing a signal over a redundant dictionary improves the compression efficiency, especially at low bit rates where most of the signal energy is captured by only few elements. Moreover, signals resulting from decomposition over redundant dictionaries are more robust to noise [1]. The main limitation of non-orthogonal transforms is however the encoding complexity, since the number of possible decompositions becomes infinite. Matching Pursuit algorithms [2] provide an interesting way to iteratively decompose the signal in its most important features with a limited complexity. It outputs a stream composed of atoms or basis functions along with their respective coefficients.

The aim of this paper is to study the effects of coefficients quantization onto reconstruction of Matching Pursuit streams. Since the Matching Pursuit coefficients generally take on real values, quantization is necessary to reduce the bandwidth needed to transmit them. Quantization errors have been studied in [3], [4] in the context of overcomplete frame expansions and consistent Matching Pursuit. This paper studies *a posteriori* coefficient quantization for a general Matching Pursuit decomposition, in contrary to usual schemes [5], [6], [7] where the encoder uses the quantized coefficients to update the residual signal. It therefore targets particularly the transmission of Matching Pursuit streams over heterogeneous networks where the stream is computed once and then quantized several times to satisfy different rate constraints.

The coefficients energy can be upper-bounded by a exponential decay curve along the iteration number. This curve only depends on the properties of the dictionary and the search algorithm. Hence, the contribution of each Matching Pursuit elements to the quantization error clearly depends on their position within the encoded stream, and hence on the redundancy of the dictionary. Based on the characterization of the energy decay curve, an exponentially bounded quantization scheme is proposed for Matching Pursuit coefficients. The quantization range and the number of quantization steps are adapted to the relative importance of the coefficients. Moreover, the optimal number of atoms to code is also given by the exponential decay curve parameters that are the redundancy factor and the signal energy. This new quantization scheme truly outperforms uniform quantization schemes, especially at low bit rates. It is shown to achieve very good results in the practical case of image compression.

The paper is organized as follows: Section II first overviews the Matching Pursuit algorithm and presents the convergence properties of the decomposition. Section III then studies the *a posteriori* quantization of Matching Pursuit coefficients and proposes an exponential quantization which takes benefit of the properties of the encoding. Experimental results are given in Sec. IV for random and Gabor dictionaries in the practical case of image coding. Finally, concluding remarks are given in Section V.

II. Matching Pursuit and Convergence Rate

In contrast to orthogonal transforms, overcomplete expansions of signals are not unique. The number of feasible decompositions is infinite, and finding the best solution under a given criteria is a NP-complete problem. In compression, one is interested in representing the signal with the smallest number of elements, that is in finding the solution with most of the energy on only a few functions. Matching Pursuit [2] is one of the sub-optimal approaches that greedily approximates the solution to this NP-complete problem.

Matching Pursuit (MP) is an adaptive algorithm that iteratively decomposes any function f in the Hilbert space $L^2(\mathbb{R})$ in a possibly redundant dictionary of functions called *atoms* [2]. Let $\mathcal{D} = \{g_\gamma\}_{\gamma\in\Gamma}$ be such a dictionary with $\|g_\gamma\| = 1$ and Γ

This work has been partly supported by Hewlett-Packard.
Contact information: Fax: +41 216937600, Phone: +41 216934709, E-mail: pascal.frossard@epfl.ch

represents the set of possible indexes. The function f is first decomposed as follows :

$$f = \langle g_{\gamma_0} | f \rangle \, g_{\gamma_0} + \mathcal{R} f \,, \tag{1}$$

where $\langle g_{\gamma_0} | f \rangle \, g_{\gamma_0}$ represents the projection of f onto g_{γ_0} and $\mathcal{R}f$ is a residual component. Since all elements in $\mathcal{D}$ have by definition a unit norm, it is easy to see from eq. (1) that g_{γ_0} is orthogonal to $\mathcal{R}f$, and this leads to

$$\|f\|^2 = |\langle g_{\gamma_0} | f \rangle|^2 + \|\mathcal{R}f\|^2 \,. \tag{2}$$

To minimize $\|\mathcal{R}f\|$, one must choose g_{γ_0} such that the projection coefficient $|\langle g_{\gamma_0} | f \rangle|$ is maximum. The pursuit is carried out by applying iteratively the same strategy to the residual component. After N iterations, one has the following decomposition for f :

$$f = \sum_{n=0}^{N-1} \langle g_{\gamma_n} | \mathcal{R}^n f \rangle g_{\gamma_n} + \mathcal{R}^N f \,, \tag{3}$$

where $\mathcal{R}^N$ is the residual of the N^{th} step with $\mathcal{R}^0 f = f$. Similarly, the energy $\|f\|^2$ is decomposed into :

$$\|f\|^2 = \sum_{n=0}^{N-1} |\langle g_{\gamma_n} | \mathcal{R}^n f \rangle|^2 + \|\mathcal{R}^N f\|^2 \,. \tag{4}$$

Although Matching Pursuit places very few restrictions on the dictionary, the latter is strongly related to convergence speed and thus to coding efficiency. Any collection of arbitrarily sized and shaped functions can be used as dictionary, as long as completeness is respected. The completeness property ensures that Matching Pursuit is able to perfectly recover the input signal after a possibly infinite number of iterations.

The convergence speed of Matching Pursuit corresponds to its ability to extract the maximum signal energy in a few iterations. In other words, it corresponds to the decay rate of the residue and thus the coding efficiency of Matching Pursuit. The approximation error decay rate in Matching Pursuit have been shown to be bounded by an exponential [2], [8]. From [9], there exists $\alpha > 0$ and $\beta > 0$ such that for all $m \geq 0$:

$$\|\mathcal{R}^{m+1} f\| \leq (1 - \alpha^2 \, \beta^2)^{\frac{1}{2}} \, \|\mathcal{R}^m f\| \,, \tag{5}$$

where $\alpha \in (0, 1]$ is an optimality factor. This factor depends on the algorithm that, at each iteration, searches for the best atom in the dictionary. The optimality factor α is set to one when MP browses the complete dictionary at each iteration. The parameter β depends on the dictionary construction. It represents the ability of the dictionary functions to capture features of any input function f and satisfies :

$$\sup_{\gamma} |\langle f, g_{\gamma_n} \rangle| \geq \beta \|f\|. \tag{6}$$

The redundancy factor β corresponds thus to the cosine of the maximum possible angle between a direction f and its closest direction among all dictionary vectors. A general formulation of the redundancy can be found in [10].

III. A Posteriori Quantization of Matching Pursuit Coefficients

A. Signal Reconstruction Error

In practical cases an exact reconstruction of the signal is rarely possible. Indeed, the MP coder generally outputs real coefficients, which clearly require quantization to limit the coding rate. *A posteriori* quantization is very useful in an asymmetric context where a stream is encoded once and then quantized several times to meet different rate constraints. The MP algorithm is indeed the heaviest part of the encoding. A different encoding for each client as needed by in-loop quantization would therefore be too demanding in this case.

Let c_{γ_n} denote a MP coefficient, or equivalently the inner product $\langle g_{\gamma_n} | \mathcal{R}^n f \rangle$. The squared reconstruction error Δ between the signal approximation f_N and its reconstructed version $\tilde{f}$ can generally be written as :

$$\Delta = \|f_N - \tilde{f}\|^2 \; = \| \sum_{n=0}^{N-1} c_{\gamma_n} \, g_{\gamma_n} - \sum_{n=0}^{N-1} \tilde{c}_{\gamma_n} \, g_{\tilde{\gamma_n}} \|^2 \,, \tag{7}$$

where $\tilde{c}_{\gamma_n}$ and $g_{\tilde{\gamma_n}}$ are respectively the distorted coefficient and atom. Let $\xi_n = c_{\gamma_n} - \tilde{c}_{\gamma_n}$ denote the error on the coefficient. It will be assumed in the remainder of this paper that atom indexes are correctly received, i.e., that $g_{\tilde{\gamma_n}} = g_{\gamma_n}$. By triangular inequality, Δ can thus be bounded as

$$\Delta = \| \sum_{n=0}^{N-1} (c_{\gamma_n} - c_{\tilde{\gamma_n}}) \, g_{\gamma_n} \|^2 \leq \sum_{n=0}^{N-1} |\xi_n|^2 \, \|g_{\gamma_n}\|^2 \leq \sum_{n=0}^{N-1} |\xi_n|^2 \,, \tag{8}$$

since $\|g_\gamma\| = 1$. Eq. (8) provides an upper-bound to the reconstruction error due to either quantization or even transmission noise. This one can thus be bounded by the sum of the error on the coefficients. Due to the exponential convergence of the MP, this error depends generally almost only on the first coefficients, i.e., the most important ones.

B. Coefficient Norm Decay

Since Matching Pursuit elects at each step only the most important atom from an overcomplete dictionary, the coefficients have a decreasing importance along the iteration number. Recall that the coefficient c_{γ_n} corresponds to the scalar product $|\langle g_{\gamma_n}|\mathcal{R}^n f\rangle$. A Matching Pursuit iteration can be thus characterized as

$$\mathcal{R}^n f = c_{\gamma_n} g_{\gamma_n} + \mathcal{R}^{n+1} f. \tag{9}$$

Since $\mathcal{R}^{n+1} f$ is orthogonal to g_{γ_n} thanks to Matching Pursuit properties, an energy conservation relation could be written for each iteration:

$$\|\mathcal{R}^n f\|^2 = |c_{\gamma_n}|^2 + \|\mathcal{R}^{n+1} f\|^2, \tag{10}$$

where $\|g_{\gamma_n}\| = 1$. Eq. (10) means that the norm of the coefficient is strongly related to the energy of the residue. From eq. (5), this norm is thus upper bounded by an exponential function, which can be written as

$$|c_{\gamma_n}| \leq (1-\alpha^2\,\beta^2)^{\frac{n}{2}}\,\|f\|\,. \tag{11}$$

The norm of the coefficient is therefore bounded by an exponential function which depends on both the energy of the input function and the construction of the dictionary. More particularly, highly redundant dictionaries induce a very rapid decay of the coefficients norm. Indeed, since the coefficient can obviously not bring more energy than the residual function, the norm of the coefficient is strongly related to the residual energy decay curve.

C. Redundancy-Driven Uniform Coefficients Quantization

The aim of the quantization is obviously to offer the best possible reconstruction quality for a given bit budget. The optimal solution minimizes the reconstruction error given by eq. (8). We propose a sub-optimal solution using the exponential decay of the residual energy. This interesting property directly drives the quantization of the coefficients in two ways. Intuitively, the quantization error on a Matching Pursuit element depends first on the iteration. Therefore, the highest iteration elements can be more coarsely quantized than the first elements. Second, the number of Matching Pursuit elements can also be adapted to the available bandwidth, since the highest iterations carry very low energy.

Hence, the exponential upper-bound on the coefficients is used to design an efficient quantization scheme. Clearly there is no need to quantify all coefficients on the same range, since their values decreases exponentially. Bits can thus be saved by simply limiting the quantization region between 0 and the exponential decay curve given by the parameters (i.e., β and $\|f\|$). An additional bit of sign then suffices to completely characterize the coefficients[1]. The number of coefficients, as well as the number of bits per coefficient have now to be optimized in this context of exponentially upper-bounded quantization range.

The distortion at decoding can be bounded by the sum of the quantization error and the approximation error due to the limit on the number of decoded iterations. In the case where atom indexes are correctly received, using eq. (8) the total distortion can be written as :

$$D \leq \sum_{n=0}^{N-1} |\xi_n|^2 + \|\mathcal{R}^N f\|^2 \leq \sum_{n=0}^{N-1} |\xi_n|^2 + (1\;-\;\alpha^2\,\beta^2)^N \|f\|^2\;, \tag{12}$$

where the energy of the residue at iteration N is bounded thanks to eq. (5). Assume now that the distribution of the coefficients norm is uniform between 0 and the exponential upper-bound given by eq. (11). This distribution depends on the Matching Pursuit search algorithm. Keep in mind though that the assumption of a uniform distribution is valid only in a first approximation.

For complexity reasons, and under the previous assumptions, the coefficient c_j is uniformly quantized within the exponentially decaying quantization range $I_j = \nu^j \|f\|$, where $\nu = (1\;-\;\alpha^2\,\beta^2)^{\frac{1}{2}}$. Let n_j be the number of quantization steps within I_j for the quantization of the j^{th} coefficient. In the case of uniform quantization, the distortion can therefore be written as :

$$D_Q = \sum_{n=0}^{N-1} |\xi_n|^2 = \sum_{j=0}^{N-1} \frac{\nu^{2j}\;\|f\|^2}{12\;n_j^2}\;. \tag{13}$$

The optimal quantization now minimizes the distortion for a bit rate R, or equivalently, minimizes the bit rate for a given distortion. In other words, we have to find the optimal parameters $n_j \geq 1$ and N that minimizes the distortion for a given rate. The Lagrangian multiplier method [12] is well suited for this kinds of constrained optimization problems. It defines a cost function $\mathcal{L}(\lambda)$ as the sum of the objective distortion function and the constraint on the rate, weighted by the Lagrangian multiplier λ. In our case the cost function may be written as :

$$\mathcal{L}(\lambda) = D + \lambda\,R = \sum_{j=0}^{N-1} \frac{\nu^{2j}\;\|f\|^2}{12\;n_j^2} + \nu^{2\;N}\|f\|^2 + \lambda\,\left(\sum_{j=0}^{N-1} \log_2(n_j) + N\;B\right)\;, \tag{14}$$

[1] The sign bits can be reported on the rotation index in the case of structured dictionary built on anti-symmetric functions [11].

where B represents the number of bits needed to code the atom index. The index size is assumed to be constant through the iterations, even if it is possible to perform an entropy coding of the index parameters [7]. The Lagrangian formulation allows to solve the hard constrained problem of finding the optimal set of n_j and N by converting it to a set of unconstrained problems, parametrized by λ.

The optimal quantization is obtained by differentiating $\mathcal{L}(\lambda)$ with respect to both n_j and N. First, solving

$$\frac{\partial \mathcal{L}(\lambda)}{\partial n_j} = 0, \forall j \tag{15}$$

for n_j positive and finite, the optimal quantization is given by

$$n_j = \sqrt{\frac{\|f\|^2 \, \nu^{2j} \, \log 2}{6 \, \lambda}} \, . \tag{16}$$

The solution of eq. (16) is indeed a minimum of the Lagangian since the second derivative is positive at this point, regardless of the value of j. Hence, the optimal quantization imposes an exponential law to the number of quantization levels :

$$\frac{n_{j+1}}{n_j} = \frac{\nu^{j+1}}{\nu^j} = (1 - \alpha^2 \, \beta^2)^{\frac{1}{2}} \, . \tag{17}$$

Interestingly, this bit distribution leads to an equivalent participation of each iteration to the total distortion. Indeed, the distortion per coefficient is equal to :

$$\frac{\lambda}{2 \, log(2)}, \; \forall n_j > 1 \, , \tag{18}$$

independently of the iteration. Notice however that the Lagrangian formulation provides only an approximation of the optimal quantization since n_j takes only integer powers of two values in practical cases.

The appropriate number of MP iterations has now to be found in order to minimize the reconstruction error for a given rate. The rate can indeed be limited by transmitting only part of the stream. Moreover, it can be seen from eq. (14) that indexes can be transmitted alone, even if no coefficient is coded ($n_j \leq 1$). Depending on the redundancy of the dictionary, indexes often carry more information than coefficients. One can therefore imagine a scheme where the coefficients are simply interpolated from the exponential decay curve, especially for high order iterations (i.e., small coefficients). The optimal number of iterations N is thus given by minimizing the Lagrangian cost function of eq. (14) where the n_j have been replaced by their optimal values from eq. (16) :

$$\mathcal{L}(\lambda) = \sum_{j=0}^{N-1} \mathcal{L}_j(\lambda) + \nu^{2\,N} \|f\|^2 \tag{19}$$

with

$$\mathcal{L}_j(\lambda) = \begin{cases} \frac{\nu^{2j} \, \|f\|^2}{12 \, n_j^2} + \lambda \, (\log_2 n_j + B) & \text{if } n_j \geq 1 \\ \frac{\nu^{2j} \, \|f\|^2}{12} + \lambda \, B & \text{otherwise.} \end{cases} \tag{20}$$

The Lagrangian is defined as a piecewise function since indexes can be transmitted without coefficients, as stated before. The break-point occurs at $j = N_B = \lfloor \frac{-\log(n_0)}{\log(\nu)} \rfloor$ (i.e., $n_j = 1$) where $n_0 = \sqrt{\frac{\|f\|^2 \, log(2)}{6 \, \lambda}}$ from Eq (16). Notice that N_B is positive only if $\lambda \leq \frac{\|f\|^2 \, log(2)}{6}$. Otherwise, the weight on the rate in the Lagrangian cost function becomes much more important than the distortion, and the best scheme would be not to transmit any coefficient. The optimal number of iterations, N_{opt}, is therefore obtained by the zeros of first derivative on both parts of the Lagrangian cost function. Thus,

$$N_{opt} = \begin{cases} N_1 & \text{if } B > -\frac{23 \, \log(\nu) + 1}{2 \, \log(2)} \\ N_2 & \text{otherwise.} \end{cases} \tag{21}$$

where :

$$N_1 = -\frac{1 + 2\,B \, \log(2) - \log(\nu) + \log\left(\frac{\|f\|^2 \, \log(2)}{6 \, \lambda}\right)}{2 \, \log(\nu)} - \frac{W_{-1}\left(3 \, 2^{3-2\,B} \, \nu \, \log(\nu) \, e^{-1}\right)}{2 \, \log(\nu)} \, , \tag{22}$$

$$N_2 = \frac{\log\left(-\frac{\lambda\ B}{2\|f\|^2\ \log(\nu)\ \left(1+\frac{1}{12\ \nu^2-12}\right)}\right)}{2\ \log(\nu)} \tag{23}$$

are the optimum values of N on both sides of the piecewise function (i.e., for respectively $N \leq N_B$ and $N > N_B$). The choice between N_1 and N_2 depends only on the relation between B and ν. Indeed, the atom index size and the decay rate of the coefficients norm influence the decision of coding an atom without coefficients. In eq. (22), $W_{-1}(x)$ represents the second branch of the Lambert W function [13]. Finally, these solutions are indeed a minimum of the Lagrangian cost function, since the second derivative is positive at $N = N_{opt}$. Details of the computation can be found in [14].

Now that the unconstrained problem of eq. (14) has been solved optimally for an arbitrary λ, the next step is to find the optimal λ_{opt} that guarantees a bit budget R_{budget}. The solutions of the optimization problem forms a convex hull of the achievable rate-distortion curve, and the bit budget constraint imposes a λ_{opt} that represents the slope of the rate-distortion function at $R = R_{budget}$. Several bisection methods have been proposed to solve this second problem [15], [16], [17]. Since in our case we can express the bit rate in terms of the Lagrangian multiplier λ, the constraint on the bit budget directly imposes an approximated value for λ_{opt} as :

$$R_{budget} = \begin{cases} \sum_{j=0}^{N_1-1} \log_2(n_j) + N_1\ B & \text{if } B > -\frac{23\ \log(\nu)+1}{2\ \log(2)} \\ \sum_{j=0}^{N_B-1} \log_2(n_j) + N_2\ B & \text{otherwise.} \end{cases} \tag{24}$$

where N_1, N_2 and n_j are functions of λ from the above equations. The approximation is due to the constraint of integer number of bits for each coefficient. The values of λ_{opt} can finally easily be computed through numerical methods.

IV. Experimental Results

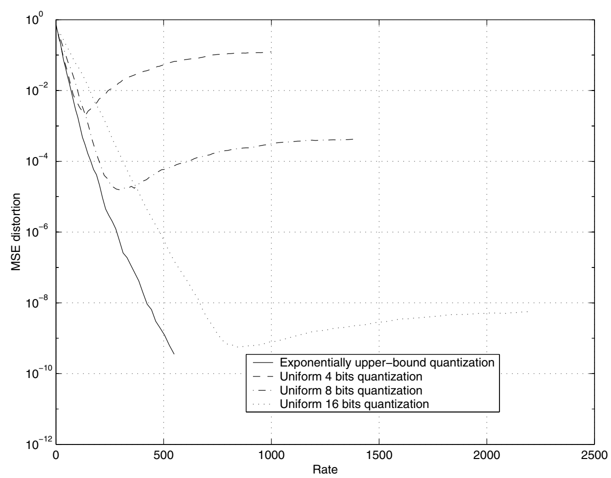

Fig. 1. R-D curve for the decomposition of a random signal of length 10 over a random dictionary of 50 vectors. Comparison of exponential quantization and uniform quantization as a function of the total rate.

This section now presents experimental results of the new quantization scheme for random and Gabor dictionaries in image coding. The exponentially upper-bounded quantization is compared to classical *a posteriori* uniform quantization methods. Fig. 1 shows the rate-distortion characteristic for a decomposition of a random signal of 10 samples over a random dictionary of 50 vectors. The redundancy-driven quantization scheme is compared to different uniform quantization schemes. The latter quantifies the coefficients over a range defined by the energy of the signal. The proposed exponentially upper-bounded quantization clearly outperforms the uniform quantization, since it adapts to the rate and to the range of coefficients to provide the best approximation for the available rate. The uniform quantization schemes provides good results for low rate, since the first coefficients are finely quantized. For high rates, however, the coefficients become too small compared to the quantization steps. The error therefore increases, since the quantized coefficient is often larger than the true value computed by the MP.

Fig. 2 shows the behavior of the redundancy-driven quantization scheme in the practical case of image coding. The *Lena* image has been decomposed onto a structured Gabor dictionary. The coefficients have then been quantized with respectively our scheme and common uniform quantization schemes. The distortion is clearly smaller for the proposed quantization scheme. However, for very low rates, uniform quantization can be slightly better due to mismatch between the coefficients decay rate and the values of the first coefficients. The mismatch is due to the sub-optimal search within the MP. Indeed, an exhaustive search would be too demanding in the case of images, for which a sub-optimal method has been provided using Genetic Algorithms [18]. This introduces a sub-optimality factor α (see eq. (5)) smaller than one and influences the decay rate of the coefficients. The reconstructed image

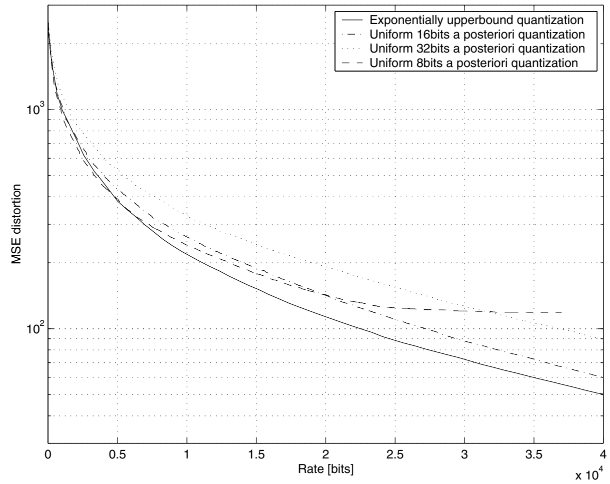

Fig. 2. Rate-distortion curves for the new quantization scheme and uniform a posteriori quantization of Lena MP stream. Distortion is represented as a function of the total rate.

are represented in Fig. 3 to 5 for similar rates. The redundancy-driven quantization provides better reconstructed images compared to *a posteriori* and even *a priori* uniform quantization. The PSNR quality is slightly better, and the visual quality is also better since more components are used for the reconstruction, for the same transmission rate. Notice that the rates given in the previous figures

Fig. 3. Reconstruction of the Lena image with Gabor atoms with exponentially upper-bounded quantization. PSNR = 26.7 dB, Rate = 16.4 kbit (without entropy coding).

Fig. 4. Reconstruction of the Lena image with Gabor atoms with a posteriori coefficients uniform 16-bits quantization. PSNR = 25.8 dB, Rate = 16.4 kbit (without entropy coding).

Fig. 5. Reconstruction of the Lena image with Gabor atoms with a priori coefficients uniform 16-bits quantization. PSNR = 25.7 dB, Rate = 16.4 kbit (without entropy coding).

represents only qualitative values, and can not be used to compared Matching Pursuit with other coding schemes. The efficiency of the MP is described for example in [19], [20]. Indeed, in our application, the dictionary is not optimized and additional entropy coding may significantly reduce the bit rate.

The exponentially upper-bounded uniform quantization is equivalent to the division of the coefficients by an exponential quantization table factors and multiplied by a quantizer scale factor which is given by the bit budget. The optimal number of iterations depends only on the bit budget and the design of the input dictionary. Practically speaking, the only parameters to transmit for the decoder to perform the inverse quantization are ν, $\|f\|$ and n_0, or equivalently the quantizer scale factor.

Finally, the redundancy-driven quantization becomes particularly interesting for highly redundant dictionary, where the values of the coefficients decreases very rapidly. In this case, the distribution of bits among coefficients is particularly efficient compared to uniform quantization. Similarly, for small index size the proposed scheme is also more efficient since additional elements can be sent at low price.

V. Conclusions

A new approach of Matching Pursuit coefficients coding has been proposed in this paper. It describes an *a posteriori* quantization of the Matching Pursuit stream to meet the constraints of asymmetric applications targeting an heterogeneous set of decoders. The encoder takes advantage of the intrinsic properties of the Matching Pursuit and more particularly of the exponential decay of the coefficients to perform a redundancy-driven coefficient quantization. The quantization range as well as the number of quantization levels are reduced exponentially to distribute bits according to the coefficients importance. The number of iterations to transmit is also optimally chosen depending on the rate constraint. The new quantization scheme clearly outperforms uniform quantization schemes since it adapts to the coefficient decay to provide the best possible approximation with the lowest coding rate. It even provides better results than in-loop quantization, while offering additional design flexibility since it is completely disconnected from the encoding. Finally, entropy coding was not considered in this paper. Entropy coding can be performed on top of the quantization to reduce coding rate [7]. One can also think about an entropy-constrained quantization, where the entropy is considered in the design of the quantization scheme.

REFERENCES

[1] Cvetkovic Z., "Accurate Subband Coding with Low Resolution Quantization," in *Proceedings of the IEEE Data Compression Conference*, 1998, pp. 448–457.

[2] Mallat S.G. and Zhang Z., "Matching Pursuits With Time-Frequency Dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, December 1993.

[3] Goyal V.K., Vetterli M. and Thao N.T., "Quantized Overcomplete Expansions in R^N: Analysis, Synthesis and Algorithms," *IEEE Transactions on Information Theory*, vol. 44, no. 1, pp. 16–31, January 1998.

[4] Cvetkovic Z., "Source Coding with Quantized Redundant Expansions: Accuracy and Reconstruction," in *Proceedings of the IEEE Data Compression Conference*, 1999, pp. 344–358.

[5] Goyal V.K. and Vetterli M., "Dependent coding in quantized matching pursuit," in *Proceedings of the SPIE - Visual Communication and Image Processing*, 1997, vol. 3024, pp. 2–12.

[6] Neff R. and Zakhor A., "Adaptive Modulus Quantizer Design for Matching Pursuit Video Coding," in *Proceedings of the IEEE International Conference on Image Processing*, 1999, vol. 2, pp. 81–85.

[7] Gharavi-Aikhansari M., "A model for entropy coding in matching pursuit," in *Proceedings of the IEEE International Conference on Image Processing*, 1998, vol. 1, pp. 778–782.

[8] Davis G., Mallat S. and Avellaneda M., "Adaptive Greedy Approximations," *Journal of Constructive Approximations*, vol. 13, pp. 57–98, 1997.

[9] Mallat S., *A Wavelet Tour of Signal Processing*, Academic Press, 2 edition, 1999.

[10] Frossard P. and Vandergheynst P., "Redundancy in Non-Orthogonal Transforms," in *IEEE International Symposium on Information Theory*, Washington D.C., June 2001, submitted paper.

[11] Frossard P., Vandergheynst P. and Kunt M., "Redundancy-Driven A Posteriori Matching Pursuit Quantization," *IEEE Transactions on Signal Processing*, November 2000, submitted paper.

[12] Everett H., "Generalized Lagrange Multiplier Method for Solving Problems of Optimum Allocation of Resources," *Operations Research*, vol. 11, pp. 399–417, 1963.

[13] Corless R.M., Gonnet G.H., Hare D.E.G., Jeffrey D.J. and Knuth D.E., "On the Lambert W Function," *Advances in Computational Mathematics*, vol. 5, pp. 329–359, 1996.

[14] Frossard P., *Robust and Multiresolution Video Delivery: From H.26x to Matching Pursuit Based Technologies*, Ph.D. thesis, Swiss Federal Institute of Technology, Lausanne, Switzerland, December 2000.

[15] Schuster G.M. and Katsaggelos A.K., "An Optimal Quadtree-Based Motion Estimation and Motion-Compensated Interpolation Scheme for Video Compression," *IEEE Transactions on Image Processing*, vol. 7, no. 11, pp. 1505–1523, November 1998.

[16] Shoham Y. and Gersho A., "Efficient Bit Allocation for an Arbitrary Set of Quantizers," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 36, no. 9, pp. 1445–1453, September 1988.

[17] Ramchandran K. and Vetterli M., "Best Wavelet Packet Bases in a Rate-Distortion Sense," *IEEE Transactions on Image Processing*, vol. 2, no. 4, pp. 160–175, April 1993.

[18] Figueras R.M., "Image coding with Matching Pursuit," M.S. thesis, Swiss Federal Institute of Technology, Lausanne, Switzerland, August 2000.

[19] Al-Shaykh O.K., Miloslavsky E., Nomura T., Neff R. and Zakhor A., "Video Compression Using Matching Pursuits," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 1, pp. 123–143, February 1999.

[20] Neff R., Nomura T. and Zakhor A., "Decoder Complexity and Performance Comparison of Matching Pursuit and DCT-Based MPEG-4 Video Codecs," in *Proceedings of the IEEE International Conference on Image Processing*, Chicago, IL, October 1998, vol. 1, pp. 783–787.