SCHOOL OF ENGINEERING - STI
SIGNAL PROCESSING INSTITUTE
Rosa M. Figueras i Ventura, Pierre Vandergheynst and Pascal Frossard

CH-1015 LAUSANNE

Telephone: +4121 6932601
Telefax: +4121 6937600
e-mail: {rosa.figueras, pierre.vandergheynst, pascal.frossard}@epfl.ch

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

LOW RATE AND SCALABLE IMAGE CODING WITH REDUNDANT REPRESENTATIONS

Rosa M. Figueras i Ventura, Pierre Vandergheynst and Pascal Frossard

Swiss Federal Institute of Technology Lausanne (EPFL)

Low rate and scalable image coding with redundant representations.

Rosa M. Figueras i Ventura, Pierre Vandergheynst, Pascal Frossard

Signal Processing Institute, Swiss Federal Institute of Technology, Lausanne, Switzerland

Phone: + 41 21 693 2601, Fax: +41 21 693 7600

{rosa.figueras, pierre.vandergheynst, pascal.frossard}@epfl.ch

Abstract

This paper presents a low bit-rate image coding method based on a Matching Pursuit expansion, over a dictionary built on anisotropic refinement and rotation of contour-like atoms. New breakthroughs in image coding certainly rely in signal decomposition through non-separable basis functions. The coder proposed in this paper provides an adaptive way of representing images as a sum of two-dimensional features. It is shown to provide, at low bit-rates, competitive results with the state of the art in image compression, as represented by JPEG-2000 and SPIHT, showing that the visual quality generally is better in the Matching Pursuit scheme. The coding artifacts are less annoying than the ringing introduced by wavelets at very low bit rate, due to the similarities of the basis functions of the dictionary used with the impulsional response of the V1 cells. In addition to good compression performance at low bit rate, the new coder has the advantage of producing highly flexible scalable streams. These can easily be decoded at any spatial resolution, different from the original image, and the bitstream can be truncated at any point to match diverse bandwidth requirements. The spatial adaptivity is shown to be more flexible and less complex than transcoding operations generally applied to state of the art streams. Due to both its ability for capturing the most important parts of multidimensional signals, and a flexible stream structure, the image coder proposed in this paper represents an interesting solution to image coding in visual communication applications.

I. INTRODUCTION

Compression has undoubtedly reached important milestones in the past few years, not far from reaching the limits of algorithms based on orthogonal linear transforms. Wavelet- or DCT-based schemes have achieved big compression ratios due to the huge research work that has been performed in efficiently coding the transform coefficients and parameters. Nevertheless, it can be shown that wavelet implementations, for example, are not optimal for multidimensional signals, where they fail to sparsely capture regularities or contours. New breakthroughs in image and multidimensional coding therefore rely on deep changes in the signal representation, and respectively efficient compression scheme of the transform parameters. Efficient expansions can be obtained by decomposition of the image in a sum of two-dimensional, non-separable, functions. Anisotropy and orientation of the basis functions, as well as their translation all over the image, generates a (possibly redundant) basis that composes the building blocks of the signal expansion. Recent research works in image compression have investigated the possibility of introducing rotation and/or anisotropy in the coding scheme. One of these studies [1] is based on *Curvelets*, which have orientation and a fixed anisotropy of a, a^2. Another interesting work achieves anisotropy and rotation through a prune quadtree algorithm [2]. Schemes based on *Bandelets* [3] yet represent another method to take advantage of the regularity of contours.

This paper proposes to use a Matching Pursuit algorithm to generate image representations over a redundant dictionary, which is specifically designed to capture two-dimensional features of natural images. The dictionary is built by anisotropic refinement and orientation of contour-like functions to capture edges, and isotropic Gaussian functions to represent the low frequency components. Among other candidates for the signal expansion algorithm, Matching Pursuit presents the advantage of generating fully progressive streams, whose energy bounds are computable. The progressive nature of the stream is a key feature in the design of adaptive visual communications applications, where scalability in general is becoming an important requirement. Interestingly, the structure of the proposed dictionary offers intrinsic spatial scalability of the bitstreams, in addition to the rate scalability provided by the progressive nature of the Matching Pursuit representation. In other words, the novel image coder described in this paper allows to reach good performance in terms of sparsity of the signal representation, due to the non-separable, two-dimensional, basis functions in the dictionary. And at the same time, the particular implementation based on Matching Pursuit offers an intrinsic scalability, beneficial in adaptive applications. The performance of the Matching Pursuit coder is shown to compete, at low/medium bit-rates, with state of the art algorithms like JPEG-2000, both in terms of compression and scalability, while offering a higher flexibility in the stream manipulation. The visual quality is generally in favor of the MP encoder, since coding artifacts are less annoying, thanks to the similarities of the basis functions of the dictionary with the impulsional response of the HVS V1 cells [4].

This paper is organized as follows : Section II overviews the compression of multidimensional signals, and the drawbacks of current coding scheme, and emphasizes the need for non-separable representations. Section III describes the image coder based on Matching Pursuit with a dictionary built on anisotropic refinement of two-dimensional contour-like functions. Section IV presents the performance of this coder in terms of compression, and compares it with the state of the art in image compression.

The interesting scalability features of the bit stream generated by the proposed coder are discussed in Section V. Conclusions are finally given in Section VI.

II. MULTIDIMENSIONAL IMAGE REPRESENTATIONS

A. Limits of current image representation methods

These last few years, image compression has been largely dominated by the use of wavelet based transform coding techniques. Many popular compression algorithms use wavelets at their core (SPIHT, EBCOT) and the overall success of this methodology resulted in the actual JPEG2000 standard for image compression [5]. On the conceptual point of view, there are three main different reasons for this success: (i) fast algorithms based on filter banks or on the lifting scheme, (ii) nice mathematical properties, and (iii) smart adaptive coding of the coefficients.

Efficient algorithms are, of course, of paramount importance when putting a novel technique to practice, but the overall power of wavelets for image compression really lies in the second and third items. The mathematical properties of wavelets have been well studied in the fields of Computational Harmonic Analysis (CHA) and Non-Linear Approximation Theory. Generally, the central question that both theories try to answer (at least in connection with data compression) is: given a signal, how many wavelet coefficients do I need to represent it up to a given approximation error? There is a wealth of mathematical results that precisely relate the decay of the approximation error when N coefficients are used, with the smoothness of the original signal. Modeling a signal as a piecewise smooth function, it can be shown that wavelets offer the best rate of non-linear approximation. By this we mean that approximating functions that are locally Hölder α with discontinuities, by their N biggest wavelet coefficients, one obtains an approximation error in the order of $N^{-\alpha}$ and that this is an optimal result (see [6], [7] and references therein). The key to this result is that wavelet bases yield very sparse representations of such signals, mainly because their vanishing moments *kill* polynomial parts while their multiresolution behaviour allow to localize discontinuities with few non-negligible elements. Now, practically speaking, the real question should be formulated in terms of bits: how many bits do I need to represent my data up to a given distortion? The link between both questions is not really trivial: it has to take into account both quantization and coding strategies. But very efficient wavelet coding schemes exist, and many of them actually use the structure of non-negligible wavelet coefficients accross subbands.

While this situation prevails in one dimension, it gets much more problematic in two or more dimensions mainly because of the importance of geometry. Indeed, an image can still be modeled as a piecewise smooth 2-D signal with singularities, but the latter are not point like anymore. Multi-dimensional singularities may be highly organized along embedded submanifolds and this is exactly what happens at image contours for example. Figure 1 shows that wavelets are inefficient at representing contours because they cannot deal with the geometrical regularity of the contours themselves. This is mainly due to the isotropic refinement implemented by wavelet basis : the dyadic scaling factor is applied in all directions, where clearly it should be fine along the direction of the local gradient and coarse in the orthogonal direction in order to efficiently *localize* the singularity in a sparse way.

(a) 6 coefficients   (b) 15 coefficients   (c) 25 coefficients
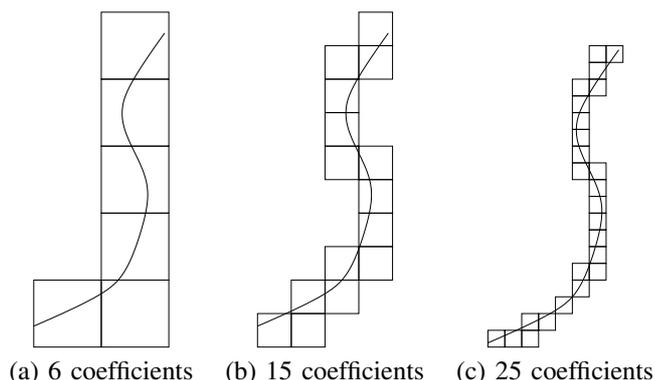
Fig. 1. Inadequacy of isotropic refinement for representing contours in images. The number of wavelets intersecting the singularity is roughly doubled when the resolution increases.

B. Anisotropic refinement

In order to efficiently represent contours, beyond the performance of wavelet decompositions, anisotropy is clearly desirable in the coding scheme. Several authors have explored the rate-distortion characteristics of anisotropic systems for representing edge-dominated images [8], [9]. These preliminary studies show that for images that are smooth away from a smooth edge (typically a $\mathcal{C}^2$ rectifiable curve), a R-D behaviour of the form

$$D(R) \asymp \frac{\log R}{R^2} \tag{1}$$

can be reached. Comparing this with the associated wavelet R-D behaviour, i.e $D(R) \asymp R^{-1}$, one clearly sees how the use of a geometry adapted system of representation can boost coding expectations. It is important to realize here that it is really the anisotropic scaling of the basis functions that allows for such performances. Simply using an anisotropic basis with multiple orientations but a fixed isotropic scaling law would not provide such results (though it may improve visual quality for instance).

Candes and Donoho [10] have recently proposed a construction called *the curvelet transform* which aims at solving the lack of flexibility of wavelets in higher dimensions. Basically curvelets satisfy an anisotropic scaling law that is adapted to representing smooth curves in images. Curvelet tight frames have been shown to achieve a much better non-linear approximation rate than wavelets for images that are smooth away from a $\mathcal{C}^2$ edge. Very interesting results have been reported for statistical estimation and denoising [11] and efficient filter bank implementations have been designed [1]. On the coding side, curvelets satisfy the localization properties that lead to (1) and there is thus hope to find efficient compression schemes based on the curvelet transform, even though such results have not yet been reported.

C. Highly non-linear approximations

Another way of achieving sparsity for low bit rate image coding is to turn to non-linear representations. In particular we will focus in this paper on the use of highly non-linear approximations in redundant dictionaries of functions.

Highly non-linear approximation theory is mainly concerned with the following question : given a collection $\mathcal{D}$ of elements of norm one in a Banach space [1] $\mathcal{H}$, find an exact N-sparse representation of any signal s :

$$s = \sum_{k=0}^{N-1} c_k g_k \,. \tag{2}$$

The equality in (2) may not need to be reached, in which case a N-term approximant $\tilde{s}_N$ is found :

$$\tilde{s}_N = \sum_{k=0}^{N-1} c_k g_k, \quad \|s - \tilde{s}_N\| \leq \epsilon(N) \,, \tag{3}$$

for some approximation error ϵ. Such an approximant is sometimes called (ϵ, N)-sparse.

The collection $\mathcal{D}$ is often called a dictionary and its elements are called atoms. There is no particular requirements concerning the dictionary, except it should span $\mathcal{H}$ and there is no prescription on how to compute the c_k's in (2).

The main advantage of this class of techniques is the complete freedom one has in designing the dictionary used, which can then be efficiently tailored to match signal structures. An efficient dictionary design is also important from the theoretical point of view, since without restrictions on $\mathcal{D}$, computing (2) is an NP hard combinatorial problem! It has to be stressed out however that these results do not impair the possibility of finding very good non-linear approximations in *particular* classes of dictionaries, as we will see later.

Our ultimate goal would be to find the best, that is the sparsest, possible representation of the signal. In other words, we would like to solve the following problem:

$$\text{minimize } \|c\|_0 \text{ subject to } s = \sum_{k=0}^{K-1} c_k g_{\gamma_k},$$

where $\|c\|_0$ is the number of nonzero entries in the sequence $\{c_k\}$. If the dictionary is well adapted to the signal, there are high hopes that this kind of representation exists, and would actually be sparser than a nonlinear wavelet-based approximation. As previously stated, the problem of finding a sparse expansion of a signal in a generic dictionary leads to a daunting NP hard combinatorial optimization problem. This is however not true anymore for *particular* classes of dictionaries. Recently, constructive results were obtained by considering incoherent dictionaries, i.e. collections of vectors that are not too far from an orthogonal basis. Starting with a dictionary obtained by merging two orthogonal basis, Donoho and Huo [13] first showed that a sufficiently sparse solution is always unique and can be exactly recovered using Basis Pursuit [14]. Their findings were later refined in [15] and extended to a broader class of incoherent dictionaries by Gribonval and Nielsen [16]. These results impose very tough constraints on the dictionary but yield a striking improvement : they allow to solve the original NP hard combinatorial problem by linear programming. As we will now see, this rigidity can be relaxed when we turn to the problem of finding sufficiently good N terms approximants instead of exact solutions.

D. Greedy algorithms: Matching Pursuit

Greedy algorithms iteratively construct an approximant by selecting the element of the dictionary that best matches the signal at each iteration. The pure greedy algorithm is known as *Matching Pursuit* [17]. Assuming that all atoms in $\mathcal{D}$ have norm one, we initialize the algorithm by setting $R_0 = s$ and we first decompose the signal as

$$R_0 = \langle g_{\gamma_0}, R_0 \rangle g_{\gamma_0} + R_1 \,.$$

---

[1]A Banach space is a complete vector space B with a norm $\|v\|$, for more information please check [12].

Clearly g_{γ_0} is orthogonal to R_1 and we thus have

$$\|R_0\|^2 = |\langle g_{\gamma_0}, R_0\rangle|^2 + \|R_1\|^2\,.$$

If we want to minimize the energy of the residual R_1 we must maximize the projection $|\langle g_{\gamma_0}, R_0\rangle|$. At the next step, we simply apply the same procedure to R_1, which yields

$$R_1 = \langle g_{\gamma_1}, R_1\rangle g_{\gamma_1} + R_2\,,$$

where g_{γ_1} maximizes $|\langle g_{\gamma_1}, R_1\rangle|$. Iterating this procedure, we thus obtain an approximant after M steps:

$$s = \sum_{m=0}^{M-1} \langle g_{\gamma_m}, R_m\rangle g_{\gamma_m} + R_M\,,$$

where the norm of the residual (approximation error) satisifies

$$\|R_M\|^2 = \|s\|^2 - \sum_{m=0}^{M-1} |\langle g_{\gamma_m}, R_m\rangle|^2\,.$$

Some variations around this algorithm are possible. An example is given by the weak greedy algorithm [18], which consists in modifying the atom selection rule by allowing to choose a slightly suboptimal candidate:

$$|\langle R_m, g_{\gamma_m}\rangle| \geqslant t_m \sup_{g\in\mathcal{D}} |\langle R_m, g\rangle|\,, \quad t_m \leqslant 1\,.$$

One can easily show that Matching Pursuit converges [19] and even converges exponentially in the strong topology in finite dimension (see [17] for a proof). Unfortunately this is not true in general in infinite dimension, even though this property holds for particular dictionaries [20]. However, DeVore and Temlyakov [18] constructed a dictionary for which even a good signal, i.e., a sum of two dictionary elements, has a very bad rate of approximation: $\|s - s_M\| \geqslant CM^{-1/2}$. In this case a very sparse representation of the signal exists, but the algorithm dramatically fails to recover it! Notice though, as already stated before, that this does in no way rule out the existence of *particular* classes of very good dictionaries.

A clear drawback of the pure greedy algorithm is that the expansion of s on the linear span of the selected atoms is not the best possible one, since it is not an orthogonal projection. Orthogonal Matching Pursuit [21], [22] solves this problem by recursively orthogonalizing the set of selected atoms using a Gram-Schmidt procedure. The best M-term approximation on the set of selected atoms is thus computed and the algorithm can be shown to converge in a finite number of steps, but at the expense of a much bigger computational complexity.

Greedy algorithms offer constructive procedures for computing highly non-linear N-term approximations. Although the mathematical analysis of their approximation properties is complicated by their nonlinear nature, interesting results are emerging (see for example [23]). The choice of a particular algorithm generally consists in trading off complexity and optimality, or more generally efficiency. The new image compression scheme presented in this paper proposes to use Matching Pursuit as a suboptimal algorithm to obtain a sparse signal expansion, yet an efficient way to produce a progressive low bit-rate image representation with a controlled complexity. Matching Pursuit, as already stressed before, iteratively chooses the best matching terms in a dictionary. Despite its possible numerical complexity, MP is very easy to implement. Moreover, since there is almost no constraint on the dictionary itself, MP clearly stands as a natural candidate to implement an efficient coding scheme based on anisotropic refinement, and such a construction is detailed in the next section.

III. MATCHING PURSUIT IMAGE CODER

A. Overview
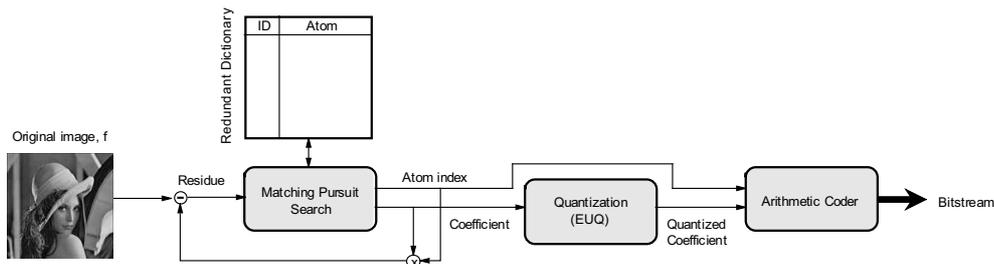
Fig. 2. Block diagram of the Matching Pursuit image coder.

The compression scheme proposed in this paper is represented in Figure 2. The image is first recursively decomposed, by Matching Pursuit, in a series of atoms chosen from a redundant dictionary, with their respective coefficients. The coefficients

are then quantized by means of an exponentially bounded uniform quantization method adapted to progressive Matching Pursuit stream characteristics (see Sec. III-C for a brief description). Coefficients and atom indexes are finally entropy coded with a context adaptive arithmetic coder. Quantization and coding are performed *a posteriori*, after the Matching Pursuit expansion. This choice is justified by the fact that the Matching Pursuit algorithm clearly represents the most computationally expensive part of the encoder. If the quantization is performed *a priori* (i.e., in the MP loop), the generated expansion is specifically destined to an target streaming rate. *A posteriori* quantization and coding allow for one single expansion to be encoded at different target rates. The image is finally reconstructed by performing the reverse operations at the decoder.

One of the first papers that proposed to use Matching Pursuit for representing images is [24]. However, it has not proposed a coder implementation, based on the MP expansion. Also, it does not use anisotropic refinement, which really provides a large benefit in representing edges and contours. Matching Pursuit has been used for coding the motion estimation error in a video sequence [25], in a block-based implementation. This coder, contrarily to ours, makes use of subblocks, which, in a sense, limits the efficiency of the expansion. In the same time, it has been designed to code the residual error of motion estimation, which presents very different characteristics than edge-dominated natural images, that represent the target application of the novel coder presented here.

One of the well-known drawbacks of Matching Pursuit is the complexity of the search algorithm, which has to be repeated at each iteration. The problem can be alleviated by sup-optimal implementations, based for example on Genetic Algorithms [26], and by heuristics linked to the dictionary. Such methods greatly speed up the search, but often sacrifice in the quality of the approximation. They sometimes get trapped in local minima, and choose atoms which do not always maximize the projection coefficient $|\langle g_\gamma | f \rangle|$. Other solutions can be found in efficient implementations of the Matching Pursuit algorithm, taking benefit from the structure of the signal and the dictionary, that might for example be decomposed in incoherent blocks. The actual implementation of the proposed encoder performs a full search over the complete dictionary, and computes all the projections in the Fourier domain [27]. This tremendously reduces the number of computations in the particular case of our dictionary built on anisotropic refinement of rotated atoms, as described here-below. The number of multiplications in this case only depends on the number of scales and rotations.

B. Anisotropic refinement using Matching Pursuit

1) Generating functions: The dictionary used in this paper is built by varying the parameters of a basis function, thus generating an overcomplete set of functions spanning the input image space. The choice of the generating function, g, is driven by the idea of efficiently approximating contour-like singularities in 2-D. To achieve this goal, the atom is a smooth low resolution function in the direction of the contour and behaves like a wavelet in the orthogonal (singular) direction. In other words, the dictionary is composed of atoms that are built on Gaussian functions along the first direction and second derivative of Gaussian functions in the orthogonal direction, that is :

$$g(\vec{p}) = \frac{2}{\sqrt{3\pi}} (4\ x^2 - 2)\ \exp(-(x^2 + y^2))\ , \tag{4}$$

where $\vec{p} = [x, y]$ is the vector of the image coordinates, and $|g| = 1$. The choice of the Gaussian envelope is motivated by the optimal joint spatial and frequency localization of this kernel. The second derivative occurring in the oscillatory component is a trade-off between the number of vanishing moments used to filter out smooth polynomial parts and ringing-like artifacts that may occur after strong quantization. It is also motivated by the presence of second derivative-like filtering in the early stages of the human visual system [28].

The generating functions described above are however not able to efficiently represent the low frequency characteristics of the image. There are two main options to capture these features: (i) to perform a low-pass filter of the image and send a quantized and downsampled image or (ii) to use an additional dictionary capable of representing the low frequency components. In order to keep the properties of full scalability, it is more appropriate to include low-pass functions in the dictionary design. This second approach has also the advantage of introducing more *natural* artifacts at very low bit rate, since it tends to naturally distribute the available bits between the low and high frequencies of the image. A second subpart of the proposed dictionary is therefore formed by Gaussian functions, in order to keep the optimal joint space-frequency localization. The second generating function of our dictionary can be written as :

$$g(\vec{p}) = \frac{1}{\sqrt{\pi}} \exp(-(x^2 + y^2))\,, \tag{5}$$

where the Gaussian has been multiplied by a constant in order to have $|g(\vec{p})| = 1$.

2) Anisotropy and orientation: Anisotropic refinement and orientation are then obtained by applying meaningful geometric transformations to the generating functions of unit L^2 norm, g, described here-above. These transformations can be represented by a family of unitary operators U_γ, and the dictionary is thus expressed as :

$$\mathcal{D} = \{U_\gamma g,\ \gamma \in \Gamma\}\ , \tag{6}$$

for a given set of indexes Γ. Basically this set must contain three types of operations :

- Translations $\vec{b}$, to move the atom all over the image.
- Rotations θ, to locally orient the atom along contours.
- Anisotropic scaling $\vec{a} = (a_1, a_2)$, to adapt to contour smoothness.

A possible action of U_γ on the generating atom g is thus given by :

$$U_\gamma g = \mathcal{U}(\vec{b}, \theta) D(a_1, a_2) g \tag{7}$$

where $\mathcal{U}$ is a representation of the Euclidean group,

$$\mathcal{U}(\vec{b}, \theta) g(\vec{p}) = g\left(r_{-\theta}(\vec{p} - \vec{b})\right), \tag{8}$$

r_θ is a rotation matrix, and D acts as an anisotropic dilation operator :

$$D(a_1, a_2) g(\vec{p}) = \frac{1}{\sqrt{a_1 a_2}} g\left(\frac{x}{a_1}, \frac{y}{a_2}\right). \tag{9}$$

It is easy to prove that such a dictionary is overcomplete using the fact that, when $a_1 = a_2$ one gets 2-D continuous wavelets as defined in [29]. It is also worth stressing that, avoiding rotations, the parameter space is a group studied by Bernier and Taylor [30]. The advantage of such a parametrization is that the full dictionary is invariant under translations and rotations. Most importantly, it is also invariant under isotropic scaling, e.g. $a_1 = a_2$. These properties will be exploited for spatial transcoding in the next sections.

3) Dictionary: Since the structured dictionary is built by applying geometric transformations to a generating mother function g, the atoms are therefore indexed by a string γ composed of five parameters: translation $\vec{b}$, anisotropic scaling $\vec{a}$ and rotation θ. Any atom in our dictionary can finally be expressed in the following form :

$$g_\gamma = \frac{2}{\sqrt{3\pi}} (4 \, g_1{}^2 - 2) \, \exp(-(g_1{}^2 + g_2{}^2)) , \tag{10}$$

with

$$g_1 = \frac{\cos(\theta) \, (x - b_1) + \sin(\theta) \, (y - b_2)}{a_1} , \tag{11}$$

and

$$g_2 = \frac{\cos(\theta) \, (y - b_2) - \sin(\theta) \, (x - b_1)}{a_2} . \tag{12}$$

For practical implementations all parameters in the dictionary must be discretized. For the Anisotropic Refinement (AR) Atoms sub-dictionary, the translation parameters can take any positive integer value smaller than the image dimension. The rotation parameter varies by increments of $\frac{\pi}{18}$, to ensure the overcompleteness of the dictionary. The scaling parameters are uniformly distributed on a logarithmic scale from one up to an eighth of the size of the image, with a resolution of one third of octave. The maximum scale has been chosen so that at least 99 % of the atom energy lies within the signal space when it is centered in the image. Experimentally, it has been found that this scale and rotation discretization choice represents a good compromise between the size of the dictionary, and the efficiency of the representation. One can choose a finer resolution for scales and rotation, getting generally more accuracy in the initial approximation. There is however a price to pay in terms of atom coding and search complexity. Finally, to further constrain the dictionary size, the atoms are always smaller along the second derivative of the Gaussian functions than along the Gaussian itself, thus maximizing the similarity of the dictionary elements with edges in images.

For the Gaussian (low frequency) sub-dictionary, the translation parameters vary exactly in the same way as for the AR atoms, but the scaling is isotropic and varies from $\frac{min(W,H)}{32}$ to $\frac{min(W,H)}{4}$ on a logarithmic scale with a resolution of one third of octave (W and H are image width and height respectively). The minimum scale of these atoms has been chosen to have a controlled overlap with the AR functions, that is large enough to ensure a good coverage of the signal space, but small enough to avoid destructive interactions between the low-pass and the band-pass dictionary. This overlap has been taken into account by having less 50% of the energy of the Gaussians in the frequency band taken by the AR fuctions. The biggest scale for these Gaussian atoms has been chosen so that at least 50% of the atom energy lies within the signal space when centered in the image. Lastly, due to isotropy, rotations are obviously useless for this kind of atoms. Sample atoms are shown in Fig. 3.

C. Coding scheme

1) Coefficient Quantization: The coefficients resulting from running Matching Pursuit with the previously described dictionary can take a priori any real value. A quantization step is clearly necessary to limit the size of the bit-stream. The proposed coder uses an *a posteriori* rate optimized exponential quantization. It takes benefit from the fact that the Matching Pursuit coefficient energy is upper-bounded by an exponential curve, decaying with the coefficient order. The quantization strongly relies on this property, since the exponential upper-bound directly determines the quantization range of the coefficient magnitude, while the coefficient sign is reported on a separate bit. The number of quantization steps is then computed as the

Fig. 3. Sample anisotropic atom with a rotation of 5*pi/18 radians and scales of 4 and 8 (a), Sample Gaussian function (b), and their respective transforms (c) and (d).

solution of a rate-distortion optimization problem. This quantization scheme has been described in details in [31]. It is briefly presented here for the sake of completeness.

Let the coefficient c_{γ_n} represent the scalar product $\langle g_{\gamma_n} | \mathcal{R}^n f \rangle$. It can be shown that its norm is upper-bounded by an exponential function [32], which can be written as

$$|c_{\gamma_n}| \leq (1 - \alpha^2 \ \beta^2)^{\frac{n}{2}} \|f\| \ . \tag{13}$$

where $\|f\|$ is the energy of the signal to code, β is a constant depending on the construction of the dictionary, and α is a sub-optimality factor depending on the Matching Pursuit implementation (for a full search algorithm as the one used in this paper, $\alpha = 1$). The coefficient upper-bound thus depends on both the energy of the input function and the construction of the dictionary. Since the coefficients can obviously not bring more energy than the residual function, the norm of the coefficient is strongly related to the residual energy decay curve.

Choosing the exponential upper-bound from Eq. (13) as the limit of the quantization range, it remains to determine the number of bits to be spent on each coefficient. The rate-distortion optimization problem shows that the number of quantization levels have also to follow a decaying exponential law, given by :

$$n_j = \sqrt{\frac{\|f\|^2 \ \nu^{2j} \ \log 2}{6 \ \lambda}} \ , \tag{14}$$

where n_j is the number of quantization levels for coefficient c_j, $\nu = (1 - \beta^2)^{\frac{1}{2}}$ is the dictionary redundancy factor, and λ is the Lagrangian multiplier which drives the size of the bit-stream [31].

In practice, the exponential upper-bound and the optimal bit distribution given by Eq. (14) are often difficult to compute, particularly in the practical case of large dictionaries. To overcome these limitations, we use a suboptimal but very efficient algorithm based on the previous optimal results. The key idea lies in a dynamic computation of the redundancy factor ν from the quantized data. Since this information is also available at the decoder, it will be able to perform the inverse quantization without any additional side information.

The MP coefficients are first re-ordered, and sorted in the decreasing order of their magnitude (this operation might be necessary since the MP algorithm does not guarantee a strict decay of the coefficient energy). The adaptive quantization scheme then performs as follows. Let $Q[c_k]$, $k = 1, \ldots j - 1$ denote the quantized counterparts of the $j - 1$ first coefficients. Due to the rapid decay of the magnitude, coefficient c_j is very likely to be smaller than $Q[c_{j-1}]$. It can thus be quantized in the range $[0, Q[c_{j-1}]]$. The number of quantization levels at step j is theoretically driven by the redundancy factor as given by Eq. (14). The adaptive quantization uses an estimate of the redundancy factor to compute the number of quantization levels as :

$$n_j = \tilde{\nu}_{j-1} \ n_{j-1} \ . \tag{15}$$

The estimate of the redundancy factor $\tilde{\nu}$ is recursively updated, replacing the quantization range, previously given by the exponential upper-bound in the optimal quantization scheme, with the quantized coefficient $Q[c_j]$ as :

$$\tilde{\nu}_j = \left(\frac{Q[c_j]}{\|f\|} \right)^{1/j} \ . \tag{16}$$

The quantization is completely determined by the choice of n_0, the number of bits for the first coefficient, and a positive value of N, the number of atoms in the signal expansion.

2) Rate Control: When the bit-stream has to conform to a given bit budget, the quantization scheme parameters n_0 and N are computed as follows. First, ν is estimated with Eq. (16) by training the dictionary on a large set of signals (e.g., images), encoded with the adaptive quantization algorithm. It can be shown empirically that the estimation quite rapidly tends to the asymptotic value of the redundancy factor. The estimation of ν is then used to compute λ as a function of the given bit budget

R_b which has to satisfy :

$$
\begin{aligned}
R_b &= \sum_{j=0}^{N-1} \log_2 n_j + \sum_{j=0}^{N-1} a_j \\
&= \sum_{j=0}^{N-1} \log_2 \nu^j + N \, \log_2 n_0 + N \, A \; ,
\end{aligned}
\tag{17}
$$

where a_j represents the number of bits necessary to code the parameters of atom c_j (i.e., positions, scales and rotation indexes), and $A = E[a_j]$ represents the average index size. From Eq. (14), the value of λ indeed determines the number of bits of the first coefficient n_0, and the number of atoms to be coded, N : under the condition that atoms whose coefficients are not quantized (i.e., $n_j < 2$) are not coded, $\nu^{N-1} \, n_0$ has to be at least equal to 2.

The adaptive quantization algorithm is then completely determined, and generally yields bit rates very close to the bit budget. For high bit-rates, it sometimes under-estimates the quantizer efficiency, due to the fact that the quantization range estimated from the quantized coefficient values is more accurate than the one obtained from the theoretical exponential upper-bound. In this case, the actual bit rate becomes smaller than the bit budget, which can however easily be met by coding additional coefficients.

Notice that, a loose estimation of ν or λ does not significantly affect the efficiency of the quantization algorithm, but rather displaces the resulting encoding on the working rate-distortion curve. Finally, several coefficients could be used for the computation of $\tilde{\nu}_j$ in Eq. 16. Such a modification will improve the accuracy of the estimation of quantization parameters, and avoid potential oscillatory effects.

3) Coding: The image coder proposed in this paper implements an arithmetic adaptive entropy coder. The core of this coder is based on [33], with the probability update method from [34]. As the distribution of the atom parameters (e.g., positions or scales) is dependent on the image to be coded, the entropy coder initializes the symbol probabilities to a uniform distribution. Each of the atom parameters (projection coefficient, positions, scales, etc.) are coded using independent context. In order to keep the intrinsic scalability features of the bit-stream, the parameters are sent in their natural order, along the decreasing order of the coefficient magnitude. This method results in a progressive stream, which can be cut at any point to generate rate scalable streams (see Sec. V). Scalability is the main motivation for choosing this kind of arithmetic coder, rather than more efficient coders that previously estimate the parameters distribution in order to optimally distribute the bits. Grouping atoms according to their position parameters, for example, would also increase the compression ratio. However, the stream would not be progressive anymore.

IV. CODING PERFORMANCE

A. Benefits of anisotropy

Since Matching Pursuit has already been presented as a valid alternative to wavelets or DCT in low bit rate coding schemes [25], it becomes important to show the real benefits of using an anisotropically refined dictionary to represent natural images. Anisotropy and rotation represent the core of the design of our coder, and hence clearly differentiate it from the previously proposed Matching Pursuit implementations. To show the benefits of anisotropic refinement, our dictionary has been compared to three different dictionaries, in terms of the quality of the MP expansion. The first dictionary uses oriented Gabor atoms generated by translation ($\vec{b}$), rotation (θ) and isotropic scaling (a) of a modulated gaussian function :

$$
U\big(a, \theta, \vec{b}\big) g(\vec{x}) = \frac{1}{a} \, g\big(a^{-1} r_{-\theta}(\vec{x} - \vec{b})\big) \; ,
\tag{18}
$$

with

$$
g(\vec{x}) = e^{i\vec{\omega}_0 \cdot \vec{x}} e^{-\|\vec{x}\|^2/2} \; .
\tag{19}
$$

The next dictionary is an affine Weyl-Heisenberg dictionary built by translation, dilation and modulation of the Gabor generating atom of Eq. (19) :

$$
U\big(a, \vec{\omega}, \vec{b}\big) g(\vec{x}) = \frac{1}{a} \, e^{i\vec{\omega} \cdot (\vec{x} - \vec{b})} g\big(a^{-1}(\vec{x} - \vec{b})\big) \; .
\tag{20}
$$

Finally, the last dictionary is simply built on an orthogonal wavelet basis. Figure 4 (a) shows the reconstructed PSNR as a function of the number of iterations in the MP expansions using different types of dictionaries. Clearly, anisotropic refinement allows to outperform the other dictionaries, in terms of approximation rate, which corresponds to the results presented in [8], [9]. As expected, the orthogonal basis offers the lowest approximation rate, and redundancy in a carefully designed dictionary provides better low rate signal representations. This comparison also shows that the use of rotations is also of interest since the oriented Gabor dictionary gives better results than the modulated one. It is worth noticing that rotation and anisotropic scaling are true 2-D transformations: the use of non-separable dictionaries is clearly beneficial to efficiently approximate 2-D objects. Separable transforms, although they may enable faster implementations, are unable to cope with the geometry of edges.
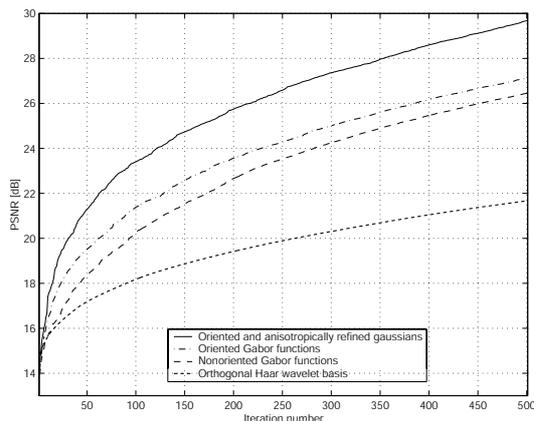
(a) Approximation rate
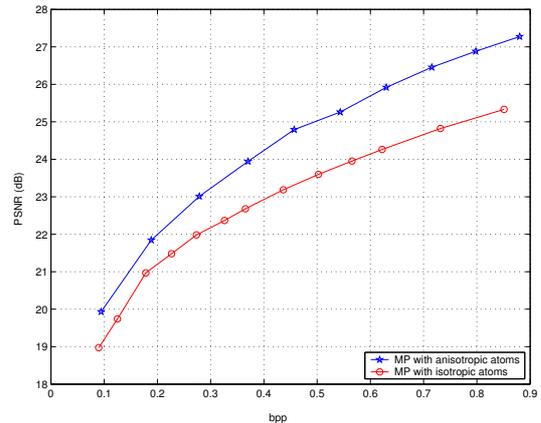

(b) Rate-distortion performance

Fig. 4. Comparison (a) of the quality of MP approximation of *Lena* (128x128 pixels), using four different dictionaries (anisotropic scaling, Gabor wavelets, Weyl-Heisenberg dictionary, and an orthogonal Haar wavelet basis) and (b) of the rate distortion of anisotropic refinement in the dictionary.

It is interesting now to analyze the penalty of anisotropy on the coding rate. In our coder, the addition of anisotropy induces the cost of coding an additional scaling parameter for each atom. To highlight the coding penalty due to anisotropic refinement, the image has also been coded with a dictionary built on isotropic atoms, all other parameters staying identical to the proposed dictionary. Figure 4 (b) illustrates the quality of the MP encoding of *Lena*, as a function of the coding rate, with both dictionaries. It is shown that the dictionary based on anisotropic refinement provides superior coding preformance. The penalty due to the coding cost of one additional scale parameter, is largely compensated by a better approximation rate. Anisotropic refinement is thus clearly an advantage in MP image coding.

B. Rate-distortion performance

Now that the benefits of anisotropic refinement have been demonstrated, the compression performance of the proposed image coder is compared to the state-of-the-art JPEG-2000 standard [2], and with the SPIHT encoder [35]. Figure 5 shows the rate-distortion performance of the three coders for common test images, at low to medium bit rates. It can be seen that Matching Pursuit provides better PSNR rating than JPEG2000 at low coding rates. However, the gap between both coding schemes rapidly decreases when the bit rate increases, as expected. Matching Pursuit and overcomplete expansions are especially efficient for low bit rate coding. They very rapidly capture the most important components of the image, but Matching Pursuit then suffers from its greedy characteristic when the rate increases. It has to be noted also that the bitstream header penalizes JPEG2000 compared to Matching Pursuit, where the syntactic information is truly minimal (at most a few bits). This penalty becomes particularly important at very low bit rate.

The performance of the proposed coder is also compared to SPIHT encoder, which introduces a minimal syntactic overhead. SPIHT almost always outperforms the proposed coder on the complete range of coding rate, and tends to perform similarly to JPEG-2000 for high rates. However, the stream generated by the SPIHT encoder is in general not easily spatially scalable, while MP and JPEG-2000 offer increased flexibility for stream adaptation.

Interestingly, all coders give equivalently poor results for highly textured images like *Barbara*. None of them is really optimized for texture coding, and the dictionary used in the proposed MP encoder is primarily designed to capture geometry in images. Finally, the proposed encoder performs reasonably well in terms of rate-distortion performance, especially at low rates. When the rate increases, the saturation of the quality can be explained by the limitations of redundant transforms for high rate approximations. Hybrid coding schemes could provide helpful solutions for high rate coding.

C. Visual comparisons

This section provides visual comparisons of images encoded with the proposed MP encoder, JPEG-2000, and SPIHT. Even if the three coding schemes perform similarly in terms of PSNR, as shown in the previous section, the coding artifacts are in general visually less annoying in the MP coding algorithm. Figure 6 compares the encoding results for the *Lena* image, at low bit rate. The decoded images are quite similar in terms of PSNR, but visual comparison seems to be in favor of the MP encoder. The artifacts introduced by MP (basically a simplification of the images) are indeed less annoying for the human observer than the ringing introduced by the wavelets in JPEG-2000.

Figure 7 presents a similar comparison between MP and JPEG-2000, but for a larger image. In this case, the PSNR rating is clearly in favor of JPEG-2000. However, the image encoded with MP is visually more pleasant than the JPEG-2000 version.

---

[2] All results have been generated with the Java implementation available at http:jj2000.epfl.ch, with default settings
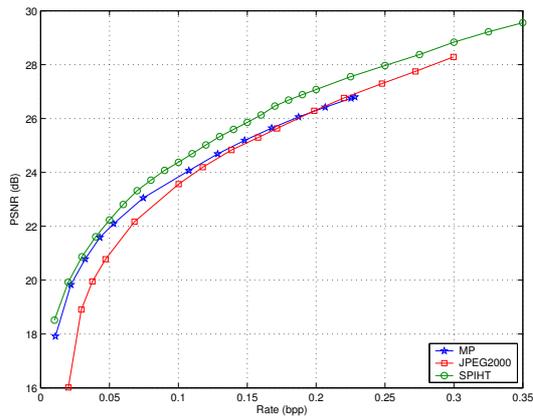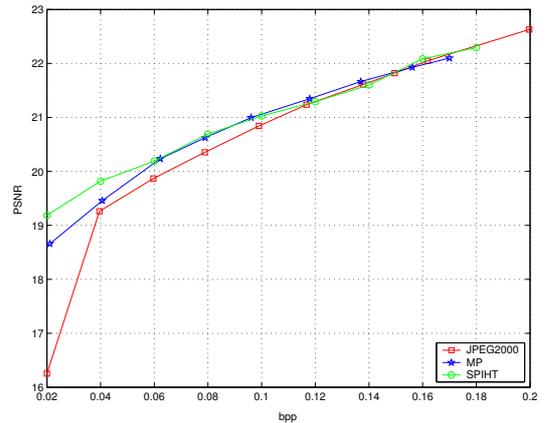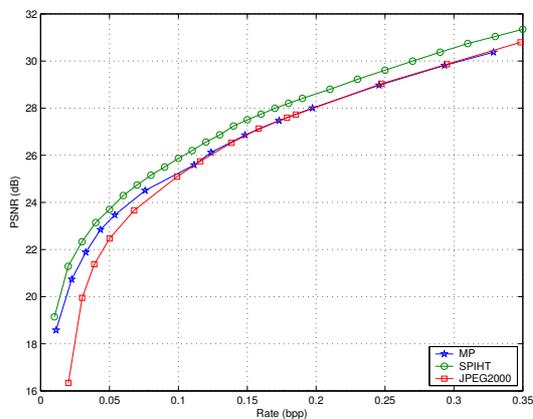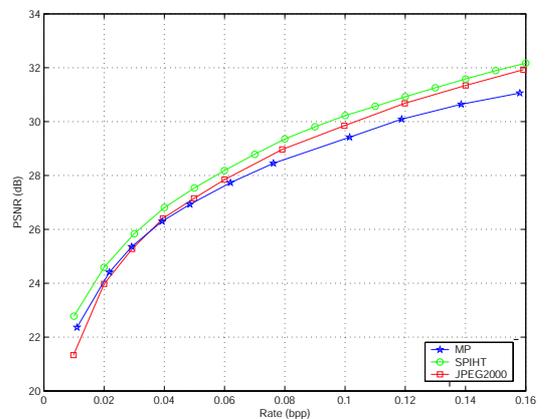
Fig. 5. Distortion-rate performance for JPEG-2000, SPIHT and the proposed MP coder, for several test images.

The detailed view of the hat, as illustrated in Figure 8, clearly confirms this impression. The JPEG-2000 encoder introduces quite a lot of ringing, while the MP encoder concentrates its effort on providing a good approximation of the geometrical patterns. JPEG-2000 has difficulties to approximate the 2-D oriented contours, which are the most predominant components of natural images.

Finally, Figure 9 proposes a comparison of the encoding of a highly textured image (*Barbara*), at a quite low bit rate. It can be seen that the three encoders perform quite poorly in this scenario, as already discussed in the previous section. The type of artifacts introduced by the encoders is however very different. In the wavelet schemes, the low resolution subbands provide a low resolution version of the image, but a lot of disturbing noise is present in textured regions. MP first concentrates on well-defined contours, but might fail in coding important semantic parts, like the eyes. Due to the flexibility of MP, one might however propose a modified coding strategy that ensures with priority a low resolution coding of all regions of interest, even if they do not necessarily represent the most energetic features.

D. Discussions

As shown by the results presented in this section, non-separable expansions are possibly the core of new breakthroughs in image compression. Anisotropic refinement and orientation allow to efficiently capture two-dimensional objects in images, and provide very good performance at low bit rate. In addition to rapidly capturing the most important features of the images, multidimensional representations generate less annoying artifacts than wavelet decompositions, which introduce some ringing at low rate.

Matching Pursuit is however just one method of signal approximation, which shows some limitations at high rate due to its greedy nature. It however allows for quite good signal representation, and presents the advantage of requiring a very low complexity decoding structure. The encoding however still requires quite a lot of computations, and is definitely much more complex than a representation over an orthogonal basis. A lot of progress have already been made in decreasing the computational complexity of the Matching Pursuit implementation. Schemes based on the division of the dictionary in incoherent blocks for example, already permit to encode images in a few seconds. MP also has the advantage of generating inherently progressive streams, which is a very interesting property for visual communication applications, as discussed in the next Section.

(a) Original

(b) MP, 30.3626 dB

(c) JPEG-2000, 30.7940 dB

(d) SPIHT, 31.3471 dB
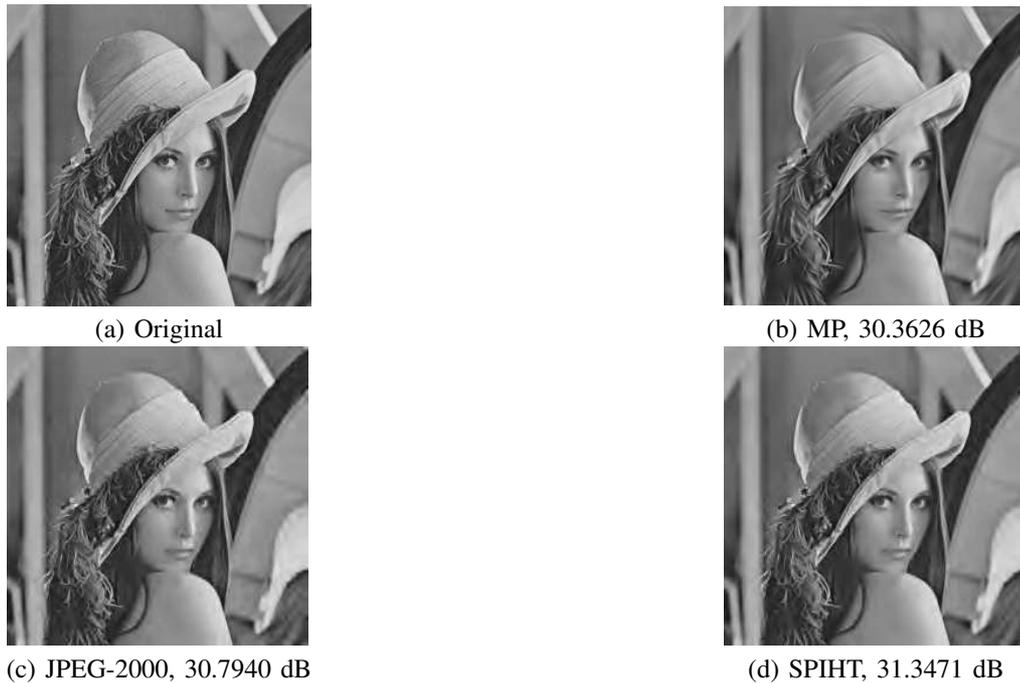
Fig. 6. *Lena* (256 x 256) encoded at 0.35bpp.

(a) MP, 31.0610dB

(b) JPEG-2000, 31.9285 dB

Fig. 7. *Lena* (512 x 512) encoded at 0.16bpp.

The coding scheme used in this paper can also surely be improved, by taking into account the proper characteristics of the image expansion. The implementation proposed in this paper has been specifically driven by the objective of building highly scalable bitstreams, but the coding efficiency can be improved when this constraint is relaxed. The PSNR rating presented in this section are generally better for JPEG-2000 than for MP. In the same time, the approximation rate in terms of number of coefficients in the expansion can be shown to be largely in favor of the Matching Pursuit scheme [36]. This clearly shows that the current coding scheme of the MP coefficients is not optimal, contrarily to the very efficient coding of wavelet coefficients in JPEG-2000. The advantage of the multidimensional decomposition in terms of approximation rate, is significantly reduced in the rate-distortion curves. A better coding scheme, adapted to the characteristics of the MP representation, is however under investigation.

V. HIGHLY FLEXIBLE SCALABILITY

A. Importance of scalability

As outlined in the previous section, one of the main advantages of the MP coder is to provide scalable streams at no additional cost. This is very interesting in nowadays visual applications involving transmission and storage, like database browsing or pervasive image and video communications. Scalability refers to the possibility for partial decoding of the stream, to fullfill

(a) MP

(b) JPEG-2000
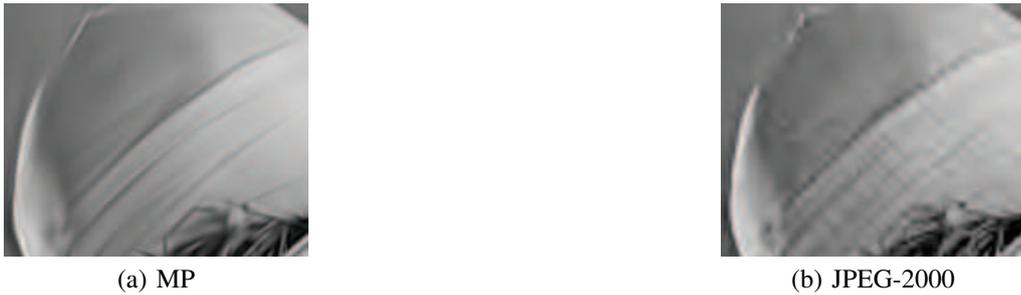
Fig. 8.   Detail view, *Lena* (512 x 512) encoded at 0.16bpp.

(a) Original

(b) MP, 21.35 dB

(c) JPEG-2000, 21.23 dB


(d) SPIHT, 21.29 dB
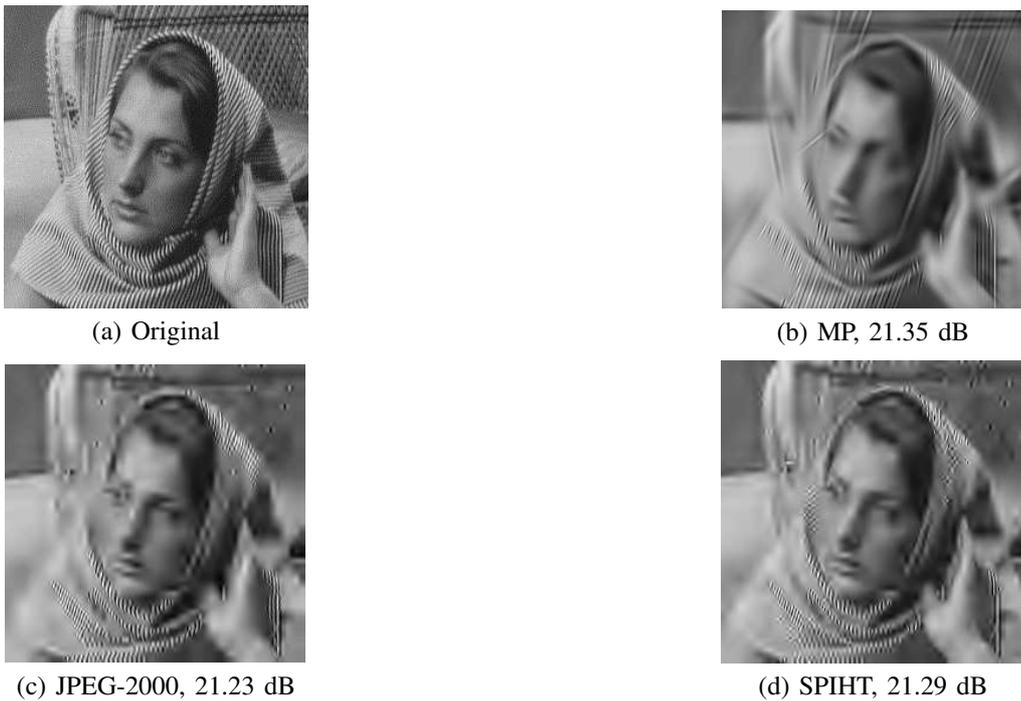
Fig. 9.   *Barbara* (256 x 256) encoded at 0.12bpp.

decoding constraints given in terms of rate, spatial resolution or complexity. The challenge in scalable coding is to build a stream decodable at different resolutions without any significant loss in quality by comparison to non-scalable streams. In other words, scalable coding is efficient if the stream does not contain data redundant to any of the target resolutions.

In image coding, scalability generally comprises rate (or SNR-) scalability and spatial scalability. On the one hand, the most efficient rate scalability is attained with progressive or embedded bitstreams, which ensure that the most important part of the information is available, independently of the number of bits used by the decoder [37], [38]. In order to enable easy rate adaptation, the most important components of the signals should be placed near the beginning of the stream. The encoding format has also to guarantee that the bitstream can be decoded, even when truncated. On the other hand, efficient spatially scalable coding schemes, like JPEG-2000 or the coder proposed in [39] are generally based on subband decompositions, which provide intrinsic multiresolution representations. However, spatial scalability is generally limited to octave-based representations, and different resolutions can only be obtained after non-trivial transcoding operations.

Multidimensional and geometry-based coding methods can advantageously provide high flexibility in the stream representation and manipulation. In this section, we will emphasize the intrinsic spatial and rate scalability of the bitstreams created with our MP image coder. First, due to the geometrical structure of the proposed dictionary, the stream can easily and efficiently be decoded at any spatial resolution. Second, the embedded bitstream generated by the Matching Pursuit coder can be adapted to any rate constraints, while the receiver is guaranteed to always get the most energetic components of the MP representation. Both types of scalability are now discussed in more details in the remainder of the section.

B. Spatial scalability

Due to the structured nature of our dictionary, the Matching Pursuit stream provides inherent spatial scalability. When both scaling parameters are equal (i.e., $a_1 = a_2$), the group law of the similitude group of $\mathbb{R}^2$ indeed applies [29] and allows for

invariance with respect to *isotropic* scaling, rotation and translation. Therefore, when the compressed image $\hat{f}$ is submitted to any combination of these transforms (denoted here by the group element η), the indexes of the MP stream are simply transformed according to the group law :

$$\mathcal{U}(\eta)\hat{f} = \sum_{n=0}^{N-1} \langle g_{\gamma_n} | \mathcal{R}^n f \rangle \, g_{\eta \circ \gamma_n} , \tag{21}$$

where

$$\left(a, \theta, \vec{b} \right) \circ \left(a', \theta', \vec{b}' \right) = \left(a \cdot a', \theta + \theta', \vec{b} + a \cdot r_\theta \vec{b}' \right) .$$

The decoder can apply the corresponding transformations to the reconstructed image simply by modifying the parameter strings of the unit-norm atoms. In other words, if η_α denotes the isotropic scaling where $\vec{a} = (\alpha, \alpha)$, the bitstream of an image of size $W \times H$ can be decoded at any resolution $\alpha W \times \alpha H$ by multiplying positions and scales by the scaling factor α. The coefficients have also to be scaled with the same factor to preserve the energy of the different components. The scaled image is thus obtained by :

$$\mathcal{U}(\eta_\alpha)\hat{f} = \alpha \sum_{n=0}^{N-1} c_{\gamma_n} g_{\eta_\alpha \circ \gamma_n} . \tag{22}$$

The transcoded atoms $g_{\eta_\alpha \circ \gamma_n}$ are given by Eqs. (10) to (12), where $\vec{b}$ and $\vec{a}$ are respectively replaced by $\alpha \vec{b}$ and $\alpha \vec{a}$. It is worth noting that the scaling factor α can take any positive real value, as long as the scaling is isotropic. Atoms that become too small after transcoding can be discarded, allowing for further bit rate reduction. Such atoms generally represent high frequency details in the image, and are located towards the end of the stream. The MP encoder sorts atoms along their decreasing order of magnitude, and scaling does not change this original arrangement. Finally, such scaling operations are quite close to image editing applications. The main difference is in the use of the scaling property. Scaling will be used at a server, or directly at the client in transcoding operations, while it could be used in the authoring tool for editing. Even in editing, the geometry-based expansion provides an important advantage over convential downsampling or interpolation functions, since there is no need for designing efficient filters. Other image editing manipulations, such as rotation of the image, or zoom in a region of interest, can easily be implemented following the same principle.

Fig. 10.   *lena* image of size 256×256 encoded with MP at 0.3bpp (center), and decoded with scaling factors of $\sqrt{\frac{1}{2}}$ (left) and $\sqrt{2}$ (right).

The simple spatial transcoding procedure is illustrated in Fig. 10, where the encoded image of size 256×256 has been re-scaled with irrational factors $\sqrt{\frac{1}{2}}$ and $\sqrt{2}$. The smallest atoms have been discarded in the down-scaled image, without impairing the reconstruction quality. The up-scaled image provides a quite good quality, even if very high-frequency characteristics are obviously missing since they are absent from the initial (compressed) bit-stream. Table I shows rate-distortion performance for spatial resizing of the 256×256 *Lena* image compressed at 0.3 bpp with the proposed Matching Pursuit coder, and JPEG-2000. It presents the PSNR values of the resized image, as well as the rate after transcoding. It also shows the PSNR values for encoding directly at the target spatial resolutions, for equivalent rates. The PSNR values have been computed with reference to the original 512×512 pixels *Lena* image, successively downsampled to 256×256 and 128×128 pixel resolutions. This is one possibility for computing the reference images and other more complex techniques, involving for example filtering and interpolation, could be adopted. They would however not significantly change the comparative results. Note that the transcoding operations for JPEG2000 are kept very simple for the sake of fairness, and the high frequency subbands are simply discarded to get the lowest resolution images.

Table I shows clearly shows that our scheme offers results competitive with respect to state-of-the-art coders like JPEG2000 for octave-based downsizing. In addition it allows for non-dyadic spatial resolutions, as well as easy up-scaling. The quality of the down-scaled images are quite similar, but the JPEG-2000 transcoded image rate is largely superior to the MP stream one. The re-scaling operation does not significantly affect the quality of the image reconstruction from MP streams. Even in the up-scaling scenario, the transcoded image provides a very good approximation of the encoding at the target (higher) resolution. In the JPEG-2000 scenario however, the adaptation of the bitstream has a quite big impact on the quality of the reconstruction, compared to an encoding at the target resolution.

Encoder		128x128	256x256	512x512
Matching Pursuit	PSNR	27.34	30.26	27.5
	Rate [bpp]	0.8	0.3	0.08
	PSNR w/o tr.	27.4	30.26	27.89
JPEG-2000	PSNR	27.18	29.99	-
	Rate [bpp]	1.03	0.3	-
	PSNR w/o tr.	33.75	29.99	-

TABLE I

COMPARISON OF SPATIAL SCALABILITY OF THE MP ENCODER AND JPEG-2000.

C. Rate scalability
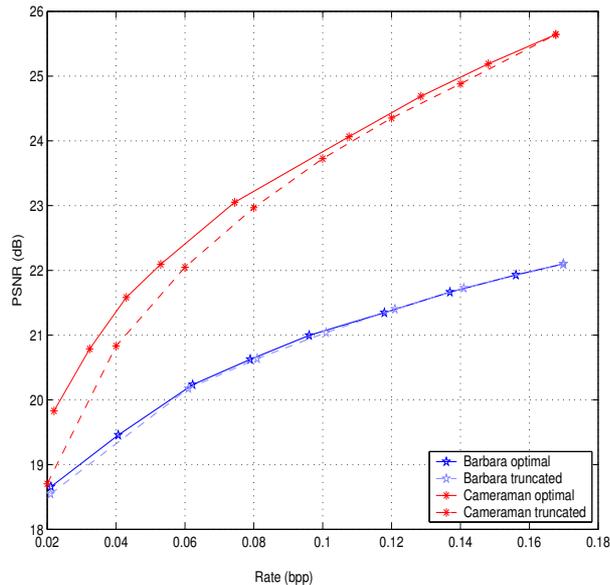
Fig. 11. Rate-distortion characteristics for MP encoding of the 256x256 *Barbara* and *Cameraman* images at 0.17 bpp, and truncation/decoding at different (smaller) bit rates.

Matching Pursuit offers an intrinsic multiresolution advantage, which can be efficiently exploited for rate adaptivity. The coefficients are exponentially decreasing so that the stream can simply be truncated at any point to provide a SNR-scalable bitstream, while ensuring that the most energetic atoms are saved. The simplest possible rate transcoding algorithm that uses the progressive nature of the Matching Pursuit stream works as follows. Assume an image has been encoded at a high target bit-rate R_b, using the rate controller described in Sec. III. The encoded stream is then restricted to lower bit budgets r_k, $k = 0, \ldots, K$ by simply dropping the bits $r_k + 1$ to R_b. This simple rate-transcoding, or filtering operation is equivalent to dropping the last iterations in the MP expansion, focusing on the highest energy atoms.

Figure 11 illustrates the rate scalability performance of the MP encoder. Images have been encoded with MP at a rate of 0.17 bpp and truncated to lower rates r_k. The bitstream has also been encoded at different target rates r_k, as described in Sec. III. It can be seen that there is a very small loss in PSNR with respect to the optimal MP stream at the same rate. This loss is due to the fact that the rate truncation simply results in dropping iterations, without using the optimal quantizer settings imposed by rates r_k as proposed in [31]. The gap is larger for images that are easier to code, since the decay of the coefficients is faster. Nevertheless, both optimal and truncated rate-distortion curves are quite close, which shows that a simple rate transcoding method, though basic, is very efficient.

VI. Conclusions

In this paper, we have presented an image coder based on a new generation of multidimensional signal representation methods. The need for image expansion as a series of non-separable, two-dimensional functions has been emphasized, that represents a way to overstep the limitations of classical representation techniques. The proposed image coder performs a Matching Pursuit decomposition of the image, with a dictionary built on anisotropic refinement of contour-like functions. The quantization scheme has been specifically designed according to the characteristics of the signal representation. The resulting encoder competes, at low bit-rates, with state of the art coders like JPEG-2000 or SPIHT, both in terms of PSNR and visual quality. The coding artifacts are in general much less annoying for the Matching Pursuit based encoder, thanks to the dictionary design.

In addition to good compression performance for low bit-rate coding, the Matching Pursuit image coding provides very interesting scalability features. Due to the structured nature of the dictionary, the bit-stream can be transcoded to any spatial resolution at a negligible computational cost. In the same time, the intrinsic multiresolution features of Matching Pursuit expansions allow for very simple rate adaptations. Extensions of this work to video coding, and particularly efficient frame-rate adaptation methods, are worth investigating as a potential solution to adaptive video delivery scenarios.

Finally, the encoder proposed in this paper has been specifically designed in order to produce scalable and easily manageable bitstreams. By relaxing these requirements, the coding performance can certainly be largely improved. It can be shown that the proposed representation clearly provides a better image representation than non-multidimensional decompositions, like wavelet-based schemes. However, this primary advantage is however compensated by a very efficient coding strategy of wavelet coefficients. A new coding scheme adapted to the proper nature of the MP decomposition is currently under study.

VII. Acknowledgments

References

[1] Do M.N. and Vetterli M., "Contourlets: A directional multiresolution image representation," in *Proceedings of the IEEE International Conference on Image Processing*, Rochester, September 2002, vol. 1, pp. 357–360.

[2] Shukla R., Dragotti P. L., Do, M. and Vetterli, M., "Rate distortion optimized tree structured compression algorithms," *IEEE Transactions on Image Processing*, 2003, submitted paper.

[3] Le Pennec E. and Mallat S., "Bandelet representation for image compression," in *Proceedings of the IEEE International Conference on Image Processing*, September 2001, pp. 12–15.

[4] Olshausen B. A. and Field D. J., "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, , no. 381, pp. 607–9, 1996, Springer-Verlag New YORK INC.

[5] Taubman D. and Marcellin M., *JPEG2000: Image Compression Fundamentals, Standards and Practice*, Kluwer Academic Publishers, Boston, November 2001.

[6] DeVore R. A., "Nonlinear approximation," *Acta Numerica*, vol. 7, pp. 51–150, 1998.

[7] Donoho D.L., Vetterli M., DeVore R. A. and Daubechies I., "Data compression and harmonic analysis," *IEEE Transactions on Information Theory*, vol. 44, pp. 391–432, 1998.

[8] Do M. N., Dragotti P.L., Shukla R. and Vetterli M., "On the compression of two dimensional piecewise smooth functions," in *IEEE International Conference on Image Processing (ICIP)*, 2001.

[9] Figueras i Ventura R. M., Granai L. and Vandergheynst P., "R-D analysis of adaptive edge representations," in *Proceedings of IEEE Intl. Workshop on Multimedia Signal Processing (MMSP02)*, St Thomas, US Virgin Islands, 2002.

[10] Candès E. J. and Donoho D. L., "Curvelets – a surprisingly effective nonadaptive representation for objects with edges," in *Curves and Surfaces*, L.L. *et al.* Schumaker, Ed., (Vanderbilt University Press, Nashville, TN, 1999.

[11] Starck J.-L., Candès E. J. and Donoho D. L., "The curvelet transform for image denoising," *IEEE Transactions on Image Processing*, vol. 11, pp. 670–684, 2002.

[12] W. Rudin, *Real and Complex Analysis*, Mc Graw Hill, 1987.

[13] Donoho D. L. and Huo X., "Uncertainty principles and ideal atomic decompositions," *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 2845–2862, November 2001.

[14] Chen S., Donoho D. L. and Saunders M. A., "Atomic decomposition by basis pursuit," *SIAM J. Scientific Comp.*, vol. 20, pp. 33–61, 1999.

[15] Elad M. and Bruckstein A.M., "A generalized uncertainty principle and sparse representations in pairs of bases," *IEEE Transactions on Information Theory*, vol. 48, no. 9, pp. 2558–2567, September 2002.

[16] Gribonval R. and Nielsen M., "Sparse representations in unions of bases," Tech. Rep. 1499, IRISA, Rennes (France), 2003.

[17] 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.

[18] DeVore R. and Temlyakov V. N., "Some remarks on greedy algorithms," *Adv. Comp. Math.*, vol. 5, pp. 173–187, 1996.

[19] Jones L., "On a conjecture of huber concerning the convergence of projection pursuit regression," *The Annals of Statistics*, vol. 15, pp. 880–882, 1987.

[20] Villemoes L. F., "Nonlinear approximation with walsh atoms," in *Surface fitting and multiresolution methods*, Schumaker Mhaut, Rabut, Ed., 1997.

[21] Pati Y. C., Rezaifar R. and Krishnaprasad P. S., "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decompositions," in *Proc. 27th Asilomar Conf. on Signals, Systems and Comput.*, 1993.

[22] Davis G. M., Mallat S. and Zhang, Z., "Adaptive time-frequency decompositions," *SPIE J. of Opt. Eng.*, pp. 2183–2191, 1994.

[23] Gilbert A. C., Muthukrishnan S. and Strauss M. J., "Approximation of functions over redundant dictionaries using coherence," in *Proc. 14th Annual ACM-SIAM Symposium on Discrete Algorithms*, 2003.

[24] Bergeaud F. and Mallat S., "Matching pursuit of images," in *Proc. IEEE Int. Conf. Image Proc.*, Washington DC, October 1995, vol. I, pp. 53–56.

[25] Neff R. and Zakhor A., "Very Low Bit-Rate Video Coding Based on Matching Pursuits," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, no. 1, pp. 158–171, February 1997.

[26] Figueras i Ventura R. M., Frossard P. and Vandergheynst P., "Evolutionary multiresolution matching pursuit and its relations with the human visual system," in *Proceedings of the European Signal Processing Conference*, Toulouse, France, September 2002, vol. 2, pp. 395–398.

[27] Figueras i Ventura R. M., Divorra Escoda O. and Vandergheynst P., "Matching pursuit full search algorithm," Tech. Rep., F-Group (LTS-2), ITS, EPFL, 2003.

[28] Marr D., *Vision*, Freeman, San Francisco, 1982.

[29] Antoine J.-P., Murenzi R. and Vandergheynst P., "Directional wavelets revisited: Cauchy wavelets and symmetry detection in patterns," *Applied and Computational Harmonic Analysis*, vol. 6, pp. 314–345, 1999.

[30] Bernier D. and Taylor K., "Wavelets from square-integrable representations," *SIAM Journal on Mathematical Analysis*, vol. 27, no. 2, pp. 594–608, 1996.

[31] Frossard P., Vandergheynst P., Figueras i Ventura R. M. and Kunt M., "A Posteriori Quantization of Progressive Matching Pursuit Streams," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 525–535, February 2004.

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

[33] Witten I.H., Neal R.M. and Cleary J.G., "Arithmetic Coding for Data Compression," *Communications of the ACM*, vol. 30, no. 6, pp. 520–540, June 1987.

[34] Duttweiler D.L.and Chamzas C., "Probability estimation in arithmetic and adaptive-huffman entropy coders," *Image Processing, IEEE Transactions on*, vol. 4 Issue: 3, pp. 237–246, Mar 1995.

[35] Said A. and Pearlman W.A., "Reversible image compression via multiresolution representation and predictive coding," in *Proceedings of the SPIE - Visual Communication and Image Processing*, 1993, vol. 2094, pp. 664–674, http://www.cipr.rpi.edu/research/SPIHT/.

[36] Vandergheynst P. and Frossard P., "Efficient image representation by anisotropic refinement in Matching Pursuit," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Salt Lake City, UT, May 2001, vol. 3, pp. 1757–1760.

[37] Said A. and Pearlman W. A., "A New, Fast, and Efficient Image Codec Based on Set Partitioning in Hierarchical Trees," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, no. 3, pp. 243–250, June 1996.

[38] Taubman D. and Zakhor A., "Multirate 3-D Subband Coding of Video," *IEEE Transactions on Image Processing*, vol. 3, no. 5, pp. 572–588, September 1994.

[39] Woods J. W. and Lilienfield G., "A Resolution and Frame-Rate Scalable Subband/Wavelet Video Coder," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 9, pp. 1035–1044, September 2001.