Low-Rate and Flexible Image Coding With Redundant Representations

Rosa M. Figueras i Ventura, Pierre Vandergheynst, *Member, IEEE*, and Pascal Frossard, *Senior Member, IEEE*

Abstract—New breakthroughs in image coding possibly lie in signal decomposition through nonseparable basis functions that can efficiently capture edge characteristics, present in natural images. The work proposed in this paper provides an adaptive way of representing images as a sum of two-dimensional features. It presents a low bit-rate image coding method based on a matching pursuit (MP) expansion, over a dictionary built on anisotropic refinement and rotation of contour-like atoms. This method is shown to provide, at low bit rates, results comparable to the state of the art in image compression, represented here by JPEG2000 and SPIHT, with generally a better visual quality in the MP scheme. The coding artifacts are less annoying than the ringing introduced by wavelets at very low bit rate, due to the smoothing performed by the basis functions used in the MP algorithm. In addition to good compression performances at low bit rates, the new coder has the advantage of producing highly flexible streams. They 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 codec bitstreams. 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 for low to medium rate image coding in visual communication applications.

Index Terms—Adaptive coding, image compression, multidimensional transform, nonlinear representation, redundant expansions.

I. INTRODUCTION

COMPRESSION has undoubtedly reached important milestones in the past few years, not far from attaining 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 orthogonal wavelet implementations, for example, are not optimal for multidimensional signals, where they fail to sparsely capture regularities or contours. Therefore, breakthroughs in image and multidimensional coding possibly 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 (2-D), nonseparable, functions. Anisotropy and orientation of the basis functions, as well as their translation all over the image, generate 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 provide orientation and a fixed anisotropy ratio of a, a^2. Another interesting work achieves anisotropy and rotation through a pruned 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 (MP) algorithm to generate image representations over a redundant dictionary, which is specifically designed to capture 2-D 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, MP 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 communication applications, where scalability in general is becoming an important requirement. Interestingly, the structure of the proposed dictionary offers intrinsic spatial adaptivity of the bitstreams, in addition to the rate scalability provided by the progressive nature of the MP 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 nonseparable, 2-D, basis functions in the dictionary. At the same time, the particular implementation based on MP offers an intrinsic flexibility, beneficial in adaptive applications. The performance of the MP coder is shown to be comparable, at low/medium bit-rates, with state of the art algorithms like JPEG2000, both in terms of compression and adaptivity. The visual quality is generally in favor of the MP encoder due to the smoothing performed by the basis functions that avoid ringing effects, and build, rather build a "sketchy" approximation of the image. That advantage might also be explained, in a first approximation, by the similarities of the basis functions of the dictionary with the impulsional response of the V1 cells of the human visual system[4],

This paper is organized as follows. Section II overviews compression of multidimensional signals and the drawbacks of current coding schemes, emphasizing the need for nonseparable representations. Section III describes the image coder based

on MP with a dictionary built on anisotropic refinement of 2-D contour-like functions. Section IV presents the performance of this coder in terms of compression, and compares it with the state of the art. The interesting adaptivity features of the bitstream are discussed in Section V. Conclusions 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 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: 1) fast algorithms based on filter banks or on the lifting scheme, 2) nice mathematical properties, and 3) 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 nonlinear approximation theory. Generally, the central question that both theories try to answer (at least in connection with data compression) is: How many wavelet coefficients do I need to represent a given signal up to a predefined 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. By modeling a signal as a piecewise smooth function, it can be shown that wavelets offer the best rate of nonlinear approximation. By this, we mean that, when 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 behavior allow to localize discontinuities with few nonnegligible 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 distribution structure of nonnegligible wavelet coefficients across 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. Multidimensional singularities may be highly organized along embedded submanifolds and this is exactly what happens at image contours, for example. Fig. 1 shows that wavelets are inefficient at representing contours because they cannot deal with their geometrical regularity. This
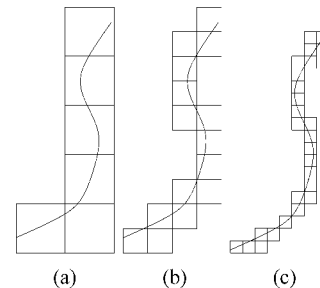

Fig. 1. Inadequacy of isotropic refinement for representing contours in images. The number of wavelets intersecting the singularity is roughly doubled when the resolution increases. (a) Six coefficients, (b) 15 coefficients, and (c) 25 coefficients.

is mainly due to the isotropic refinement of wavelet bases: The dyadic scaling factor is applied in all directions, while it rather 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.

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 behavior of the form

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

can be reached. Comparing this with the associated wavelet R-D behavior, 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 approximation performances. In particular, an anisotropic basis with multiple orientations but a fixed isotropic scaling law would not provide the same results [10] (though it may improve visual quality for instance).

Candès and Donoho [11] 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 nonlinear 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 [12] 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 hope to find efficient compression schemes based on the curvelet transform, even though such results have not yet been reported.

C. Highly Nonlinear Approximations

Another way of achieving sparsity for low bit rate image coding is to turn to nonlinear representations. In particular, we

will focus in this paper on the use of highly nonlinear approximations in redundant dictionaries of functions.

Highly nonlinear approximation theory is mainly concerned with the following: Given a collection $\mathcal{D}$ of elements of norm one in a Banach[1] space $\mathcal{H}$, find an exact N-term representation of any signal s (in the strong sense)

$$s = \sum_{k=0}^{N-1} c_k g_k. \qquad (2)$$

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

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

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 are no particular requirements concerning the dictionary, except that it should span $\mathcal{H}$, and there is no prescription on how to compute the c_ks in (2).

The main advantage of this class of techniques is the complete freedom one has in designing the dictionary, 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 nonlinear approximations in *particular* classes of dictionaries, as we will see later.

There is a clear distinction between the general framework of using redundant collections of primitives and the *actual* algorithm used to compute the approximant in (3). Indeed, using vastly redundant libraries can only improve the representation (imagine concatenating several good orthonormal bases for example). The algorithm, on the other hand, may fail to fully exploit this wealth, but, as discussed below, cases have been identified for which very well controlled algorithms do behave nicely.

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 have been obtained by considering incoherent dictionaries, i.e., collections of vectors that are not too far from an orthogonal basis.

Two classes of algorithms have been studied in this context: basis pursuit (BP) optimization strategies [13] and various flavors of greedy algorithms. Starting with a dictionary obtained by merging two orthogonal bases, Donoho and Huo [14] first showed that a sufficiently sparse solution is always unique and can be exactly recovered using BP. 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. Turning to the problem of finding sufficiently good N-term approximants instead of exact solutions, greedy heuristics offer similar advantages, as discussed in the next section.

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

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

Clearly, g_{γ_0} is orthogonal to R_1 and we 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) satisfies

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

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

$$|\langle R_m, g_{\gamma_m} \rangle| \ge t_m \sup_{g \in \mathcal{D}} |\langle R_m, g \rangle|, \qquad t_m \le 1.$$

One can easily show that MP 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 still holds for particular dictionaries [20]. 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: $\|s - s_M\| \ge CM^{-1/2}$. In this case, a very sparse representation of the signal exists, but the algorithm dramatically fails to

---

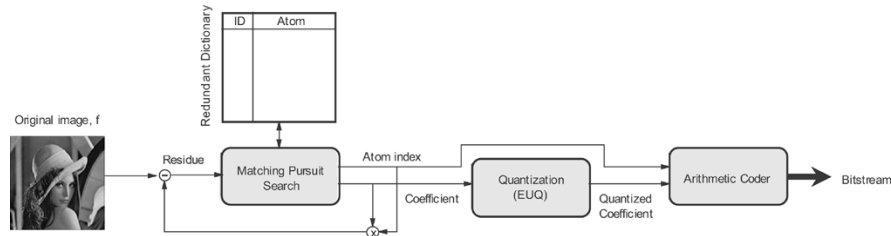[1]A Banach space is a complete vector space B with a norm $\|v\|$.

Fig. 2. Block diagram of the MP image coder.

recover it! Notice, however, that this does in no way rule out the existence of *particular* classes of very good dictionaries, even in infinite dimension, as already stated before.

A clear drawback of the pure greedy algorithm is that the expansion of s on the linear span of the selected atoms is not necessarily the sparsest one, since it is not an orthogonal projection. Orthogonal MP [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 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 nonlinear 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] and [24]). Recently, it has been shown that even the pure greedy MP algorithm can exactly recover components of a (near) optimal sparse approximation of a signal [25]. 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 MP 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. MP, as already stressed before, iteratively chooses the best matching terms in a dictionary. Despite its possible numerical complexity due to repetitive operations, 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. This construction is detailed in the next section.

III. MATCHING PURSUIT IMAGE CODER

A. Overview

The compression scheme proposed in this paper is represented in Fig. 2. The image is first recursively decomposed, by MP, in a series of atoms chosen from a redundant dictionary. Their respective coefficients are then quantized by means of an exponentially bounded uniform quantization method adapted to progressive MP stream characteristics (see Section 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 MP expansion. This choice is justified by the fact that the MP 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 a target bit 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 MP for representing images is [26]. 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, as discussed in Section II. MP has been used for coding the motion estimation error in a video sequence [27], 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, targeted by the novel coder presented here.

One of the well-known drawbacks of MP is the computational 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 [28], 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 MP algorithm, taking benefit from the structure of the signal and the dictionary, which 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 [29]. This tremendously reduces the number of computations in the particular case of our dictionary built on anisotropic refinement of rotated atoms, as described herebelow. 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, in such a way that it generates 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 one direction, and on second derivative of Gaussian functions in the orthogonal direction. The generating function reads

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

where $\vec{p} = [x, y]$ is the vector of discrete 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 [30].

The generating functions described above are, however, not able to efficiently represent the low frequency characteristics of images at low rate. There are two main options to capture these features: 1) to perform a low-pass filter of the image and send a quantized and downsampled image as side information or 2) to use an additional dictionary capable of representing the low frequency components. In order to keep the adaptivity properties, 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-frequency components of images. 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 $\|g(\vec{p})\| = 1$. It can finally be noted that another sub-dictionary could also be used for efficient texture coding. This is, however, outside the scope of this paper, which mostly concentrates on low rate coding.

2) Anisotropy and Orientation: Anisotropic refinement and orientation is obtained by applying meaningful geometric transformations g the generating functions of unit L^2 norm, described before. These transformations can be represented by a family of unitary operators $U(\gamma)$, 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 $U(\gamma)$ 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

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

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, under the restrictive condition $a_1 = a_2$, one gets 2-D continuous wavelets as defined in [10]. It is also worth stressing that, avoiding rotations, the parameter space is a group studied by Bernier and Taylor [31]. The advantage of such a parametrization is that the full dictionary is invariant under translation and rotation. Most importantly, it is also invariant under isotropic scaling, i.e., $a_1 = a_2$. These properties will be exploited for spatial adaptivity 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:

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

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 atoms sub-dictionary (AR), the translation parameters can take any positive integer value smaller than the image dimensions. The rotation parameter varies by increments of $\pi/18$, 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 scale and rotation, getting generally more accuracy in the initial steps. 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 function 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 $\min(W, H)/32$ to $\min(W, H)/4$ on a logarithmic scale with a resolution of one third of octave (W and H are image width
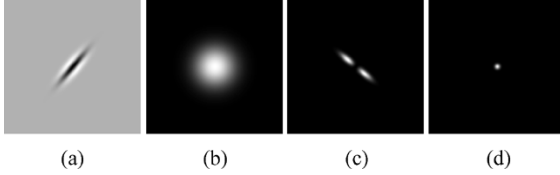
Fig. 3. (a) Sample anisotropic atom with a rotation of $5 * \text{pi}/18$ radians and scales of 4 and 8, (b) sample Gaussian function, and (c), (d) their respective transforms.

and height, respectively). The minimum scale of these atoms has been chosen to have a controlled overlap with the AR functions, i.e., 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 designed so that less than 50% of the energy of the Gaussians lies 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. Last, due to isotropy, rotations are obviously useless for this kind of atoms. Sample atoms are shown in Fig. 3.

Note that the dictionary used is a continuous dictionary, and this is the reason why the image can be reconstructed at any resolution at the decoder. In order to code 2-D discrete signals like images, the basis functions are sampled on the grip represented by discrete pixel positions $\vec{p} = [x, y]$.

C. Coding Scheme

1) Coefficient Quantization: The coefficients resulting from running MP with the previously described dictionary can take *a priori* any real value. A quantization step is clearly necessary to limit the size of the bitstream. The proposed coder uses an *a posteriori* rate optimized exponential quantization. It takes benefit from the fact that the MP coefficient energy is upper-bounded by an exponential curve, decaying with the coefficient order. The quantization strongly relies on this property: 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 solution of a rate-distortion optimization problem. This quantization scheme has been described in details in [32]. 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 [33], which can be written as

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

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 MP 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 $\mathcal{R}^n f$, the norm of the coefficient is strongly related to the residual energy decay curve.

Choosing the exponential upper-bound from (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, $\nu = (1 - \beta^2)^{1/2}$ is the dictionary redundancy factor, and λ is the Lagrangian multiplier that drives the size of the bitstream [32].

In practice, the exponential upper-bound and the optimal bit distribution given by (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 adaptive quantization scheme performs as follows. 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). 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 (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

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

Finally, the quantization process is fully 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 bitstream 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 (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

$$R_b = \sum_{j=0}^{N-1} \log_2 n_j + \sum_{j=0}^{N-1} a_j$$

$$= \sum_{j=0}^{N-1} \log_2 \nu^j + N \log_2 n_0 + N A \tag{17}$$

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 (14), the value of λ indeed determines the number of bits of the first coefficient n_0, and, subsequently, 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 rate control completely determines the adaptive quantization parameters, and generally yields bit rates very close to the bit budget. For high bit-rates, it sometimes underestimates 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 (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 [34], with the probability update method from [35]. 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. The atom parameters (projection coefficients, positions, scales, etc.) are coded using independent context. In order to keep the intrinsic scalability features of the bitstream, the parameters are sent in their natural order, along the decreasing order of the coefficient magnitudes. The bitstream is simply constructed by concatenating together coded coefficients and atom parameters. This method results in a progressive stream, which can be cut at any point to generate rate scalable streams (see Section 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, since the atoms would not anymore be ordered according to their relative importance.

IV. CODING PERFORMANCE

A. Benefits of Anisotropy

Since MP has already been presented as a valid alternative to wavelets or DCT in low bit rate coding schemes [27], 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 MP implementations. To show the benefits of anisotropic refinement, our dictionary has been compared to four different dictionaries, in terms of the quality of the MP expansion. The first dictionary uses the real part of oriented Gabor atoms generated by translation ($\vec{b}$), rotation (θ) and isotropic scaling (a) of a modulated Gaussian function , similarly to the one used in [26]

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

with

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

The next dictionary is an affine Weyl–Heisenberg dictionary built by translation, dilation and modulation of the Gabor generating atom of (19), as presented in [36]

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

where only the real part is used. The other two dictionaries are simply built on orthogonal wavelet bases. Fig. 4(a) shows the reconstructed PSNR as a function of the number of iterations in the MP expansion using different types of dictionaries. In this figure, the comparison is performed with respect to the number of terms in the expansion, in order to emphasize the approximation properties (the behavior of the coding rate is discussed below). Clearly, overcompleteness and anisotropic refinement allow to outperform the other dictionaries, in terms of approximation rate, which corresponds to the results presented in [8], [9]. As expected, the orthogonal bases offer the lowest approximation rates due to the fact that these kinds of bases cannot deal with the smoothness of edges. We can, thus, deduce that redundancy in a carefully designed dictionary provides sparser signal representations. This comparison shows, as well, that the use of rotation 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 nonseparable 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.

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 the same dictionary, built on isotropic atoms, all other parameters staying identical to the proposed scheme. Fig. 4(b) illustrates the quality of the MP encoding of *Lena*, as a function of the coding rate, with both dictionaries. To perform the comparison, the isotropic and the anisotropic dictionaries are built with the same generating function and with the same discretization of the parameters (three scales per octave and an angle resolution of $\pi/18$). The anisotropy, however, implies the coding of one additional scale parameter. It is shown that the dictionary based on anisotropic refinement provides superior coding performance, even with longer atom indexes. The penalty due to the coding cost of one additional scale parameter, is largely compensated by a better approximation rate. Anisotropic refinement is, therefore, clearly an advantage in MP image coding.
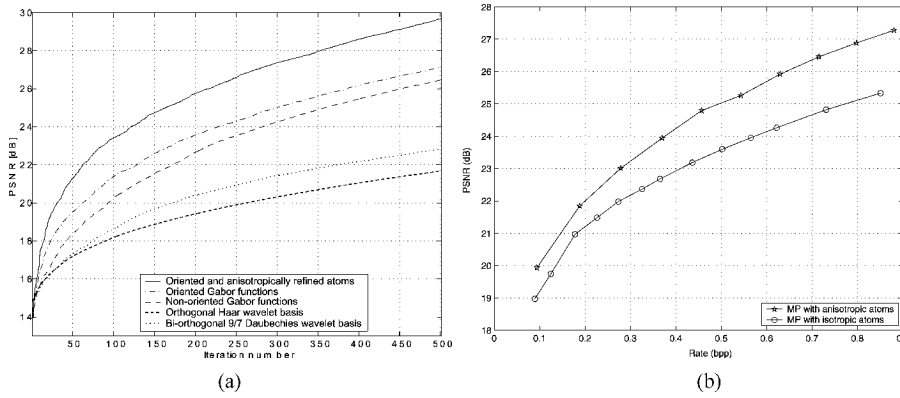
Fig. 4. Comparison of (a) the quality of MP approximation of *Lena* (128 × 128 pixels), using five different dictionaries (anisotropic refinement and oriented functions, oriented Gabor wavelets, nonoriented Gabor wavelets or Weyl-Heisenberg dictionary, an orthogonal Haar wavelet basis and a biorthogonal Daubechies 9/7 basis, with five levels of decomposition) and (b) characteristic of anisotropic refinement in the dictionary. In (b), the basis functions used are the same for isotropic and anisotropic functions, with the same parameter discretization, showing the improvements of adding anisotropy to the dictionary.
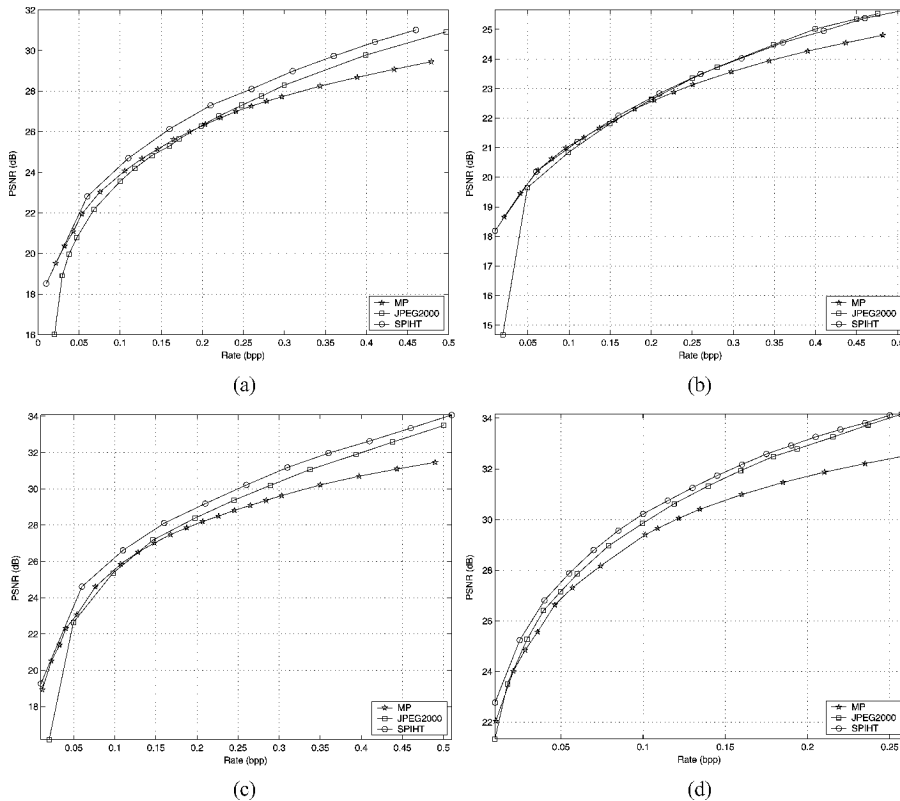

Fig. 5. Distortion-rate performance for JPEG2000, SPIHT and the proposed MP coder, for several test images. (a) *Cameraman*, 256 × 256. (b) *Barbara*, 256 × 256. (c) *Lena*, 256 × 256. (d) *Lena*, 512 × 512.

B. Rate-Distortion Performance

Now that the benefit of anisotropic refinement has been highlighted, the compression performance of the proposed image coder is compared to the state-of-the-art JPEG2000 standard,[2] and with the SPIHT encoder [37]. Fig. 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 MP 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. MP and overcomplete expansions are especially efficient for low bit rate coding. They very rapidly capture the most important components of the image, but MP then suffers from its greedy characteristic when the rate increases. Also, the dictionaries have not been designed for texture coding, but rather to code geometry. It has to be noted as well that the bitstream header penalizes JPEG2000 compared to MP, 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 the SPIHT encoder, which introduces a minimal syntactic overhead. SPIHT almost always outperforms the proposed coder on

---

[2]All results have been generated with the Java implementation available at http://jj2000.epfl.ch/, with default settings (Daubechies 9/7 wavelets and five decomposition levels).

Fig. 6. *Lena* (256 × 256) encoded at 0.35 bpp. (a) Original. (b) MP, 30.3626 dB. (c) JPEG2000, 30.7940 dB. (d) SPIHT, 31.3471 dB.

the complete range of coding rate, and tends to perform similarly to JPEG2000 for high rates. However, the stream generated by the SPIHT encoder is in general not easily spatially scalable, while MP and JPEG2000 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. Altogether, 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 coding, and by the dictionary design choices. Hybrid coding schemes could provide helpful solutions at high rate.

C. Visual Comparisons

This section provides visual comparisons of images encoded with the proposed MP encoder, JPEG2000, 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. Fig. 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 comparisons generally seem to favor the MP encoder. The artifacts introduced by MP (basically a simplification of the images) are indeed less annoying than the ringing introduced by the wavelets in JPEG2000 and in SPIHT.

Fig. 7 presents a similar comparison between MP and JPEG2000, but for a larger image. In this case, the PSNR rating is clearly in favor of JPEG2000. However, the image encoded with MP is visually more pleasant than the JPEG2000 version. The detailed view of the hat, as illustrated in Fig. 8, clearly

confirms this impression. The JPEG2000 encoder introduces quite a lot of ringing, while the MP encoder concentrates its effort on providing a good approximation of the geometrical patterns. JPEG2000 has difficulties to approximate the 2-D oriented contours, which are the predominant components of natural images. Fig. 9 further confirms the different behaviors of MP and JPEG2000. It can be seen that MP concentrates on edges, and, thus, rapidly provides a "sketchy" representation of the image. At the same rate, JPEG2000 already spends some bits in coding textures, while ringing artifacts are still visible.

Finally, Fig. 10 proposes a comparison of the encoding of a highly textured image (*Barbara*), at a quite low bit rate. As already discussed in the previous section, it can be seen that the three encoders perform quite poorly in this scenario. 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. Discussion

As shown by the results presented in this section, expansions over dictionaries built on nonseparable functions are possibly the core of new breakthroughs in image compression. Anisotropic refinement and orientation allow to efficiently capture 2-D objects in images, and provide very good performances 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.

MP 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 representations, 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 wavelet basis. A lot of progress have already been made in decreasing the computational complexity of the MP 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.

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 adaptive bitstreams, but the coding efficiency can certainly be improved when this constraint is relaxed. The PSNR ratings presented in this section are generally better for JPEG2000 than for MP. At 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 MP scheme [38]. This is partly due to the fact that

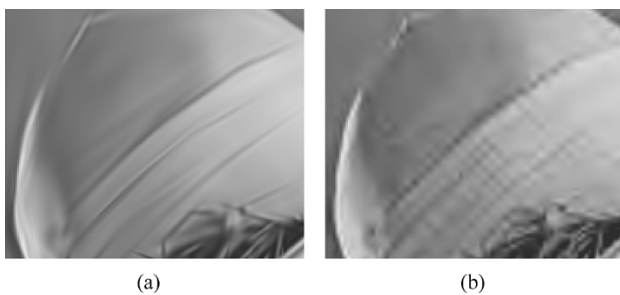Fig. 7. *Lena* (512 × 512) encoded at 0.16 bpp. (a) MP, 31.0610 dB. (b) JPEG2000, 31.9285 dB.

Fig. 8. Detail view, *Lena* (512 × 512) encoded at 0.16 bpp. (a) MP. (b) JPEG2000.

Fig. 9. *Goldhill* 256 × 256 encoded at 0.23 bpp. (a) MP, 27.49 dB. (b) JPEG2000, 28.18 dB.

Fig. 10. *Barbara* (256 × 256) encoded at 0.12 bpp. (a) Original. (b) MP, 21.35 dB. (c) JPEG2000, 21.23 dB. (d) SPIHT, 21.29 dB.

the current coding scheme of the MP coefficients is not optimal, contrarily to the very efficient coding of wavelet coefficients in JPEG2000. 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. HIGH ADAPTIVITY

A. Importance of Adaptivity

As outlined in the previous section, one of the main advantages of the MP coder is to provide highly flexible 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. We call adaptivity the possibility for partial decoding of a stream, to fulfill 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 nonadaptive streams. In other words, adaptive coding is efficient if the stream does not contain data redundant to any of the target resolutions.

In image coding, adaptivity generally comprises rate (or SNR) adaptivity and spatial adaptivity. First, the most efficient rate adaptivity 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 [39], [40]. 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. Second, schemes that implement spatial adaptivity, like JPEG2000 or the coder proposed in [41], are generally based on subband decompositions, which provide intrinsic multiresolution representations. However, spatial adaptivity is generally limited to octave-based representations, and different resolutions can only be obtained after nontrivial 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 adaptivity 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 MP coder can be adapted to any rate constraints, while the decoder is guaranteed to always receive the most energetic components of the MP representation. Most importantly, MP streams offer the advantage of decoupling spatial and rate adaptivity, which can be achieved independently. Adaptive decoding is now discussed in more details in the remainder of the section.

B. Spatial Adaptivity

Due to the structured nature of our dictionary, the MP stream provides inherent spatial adaptivity. The group law of the similitude group of $\mathbb{R}^2$ indeed applies [10] and allows for invariance with respect to *isotropic* scaling of α, rotation of Θ and translation of $\vec{\beta}$. 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 can simply be transformed with help of the group law

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

In the above expression, $\gamma_n = (\vec{a}_n, \theta_n, \vec{b}_n)$ represents the parameter strings of the atom encoded at iteration n, with scaling $\vec{a}_n$, rotation θ_n and translation $\vec{b}_n$, and $\eta = (\alpha, \Theta, \vec{\beta})$ represents the geometric transformation that is applied to the set of atoms. The decoder can apply the transformations to the encoded bitstream simply by modifying the parameter strings of the unit-norm atoms, according to the group law of similitude, where

$$(\vec{a}, \theta, \vec{b}) \circ (\alpha, \Theta, \vec{\beta}) = (\alpha \cdot \vec{a}, \theta + \Theta, \vec{b} + \alpha \cdot r_\Theta \vec{\beta}). \tag{22}$$

In other words, if $\eta_\alpha = (\alpha, 0, 0)$ denotes the isotropic scaling by a factor α, the bitstream of an image of size $W \times H$, after entropy decoding, can be used to build an image at any resolution $\alpha W \times \alpha H$ simply by multiplying positions and scales by the scaling factor α (from (22) and (7)). The coefficients have also to be scaled with the same factor to preserve the energy of the different components. The quantization error on the coefficient will, therefore, also vary proportionally to the scaling factor, but the absolute error on pixel values will remain almost unchanged,

since the atom support also varies. Finally, the scaled image is obtained by

$$\mathcal{U}(\eta_\alpha)\hat{f} = \alpha \sum_{n=0}^{N-1} c_{\gamma_n} g_{\eta_\alpha \circ \gamma_n}. \tag{23}$$

The modified atoms $g_{\eta_\alpha \circ \gamma_n}$ are simply given by (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 are discarded. This allows for further bit rate reduction, and avoids aliasing effects when $\alpha < 1$. The smallest atoms generally represent high frequency details in the image, and are located toward the end of the stream. The MP encoder initially sorts atoms along their decreasing order of magnitude, and scaling does not change this original arrangement.

Finally, 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, within intermediate network nodes, 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 conventional 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 principles. The simple spatial adaptation procedure is illustrated in Fig. 11, where the encoded image of size 256×256 has been re-scaled with irrational factors $\sqrt{1/2}$ 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) bitstream.

Table I shows rate-distortion performance for spatial resizing of the 256×256 *Lena* image compressed at 0.3 bpp, with the proposed MP coder, and JPEG2000 respectively. It presents the PSNR values of the resized image, as well as the actual bit rate after transcoding. It also shows the PSNR values for encoding directly at the target spatial resolutions, for equivalent rates. The reference images in Table I have been obtained by resizing the original 512×512 pixel *Lena* image to 256×256 and to 128×128 with a standard image processing application (Gimp). The reference image 128 DF is an image of size 128×128, obtained by taking the output of the wavelet Daubechies 9/7 filter bank at this resolution (128×128) when giving as input signal the 256×256 *Lena* reference image.

It can be seen that our scheme offers results competitive with respect to state-of-the-art coders like JPEG2000 for octave-based downsizing. The quality of the down-scaled images is slightly inferior for the MP coder, but the JPEG2000 transcoded image rate is largely superior to the MP stream one. However, MP allows for nondyadic spatial resizing, as well as easy up-scaling. The 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)

Fig. 11. *Lena* image of size 256×256 encoded with MP at (center) 0.3 bpp, and decoded with scaling factors of (left) $\sqrt{1/2}$ and (right) $\sqrt{2}$.

TABLE I
COMPARISON OF SPATIAL ADAPTIVITY OF THE MP ENCODER AND JPEG2000.
PSNR VALUES ARE COMPARED TO QUALITY OBTAINED WITHOUT
TRANSCODING (W/O TR.). 128 DF IS THE 128×128 IMAGE
OBTAINED WITH THE JPEG2000 WAVELET FILTER

Encoder		128x128	128 DF	256x256	512x512
	PSNR	28.32	24.62	29.70	27.40
MP	Rate [bpp]	0.8	0.8	0.3	0.08
	PSNR w/o tr.	29.48	28.03	29.70	28.48
	PSNR	28.90	30.53	30.19	-
JPEG2000	Rate [bpp]	1.02	1.02	0.3	-
	PSNR w/o tr.	33.66	31.08	30.19	-

resolution. Next, results are also reported for a 128×128 reference image obtained from the output of the Daubechies 9/7 filter bank, that is used in the JPEG2000 encoder. We can note that the PSNR after transcoding is obviously improved for JPEG2000 in that case, and that the PSNR value for MP is decreased. We could expect the reverse behavior with a reference image obtained with Gaussian filtering: this illustrates the importance of the reference image in spatial resizing performance evaluation. It is finally interesting to note that MP incurs a loss of almost 1.5 dB when coding that particular reference image, mostly due to the aliasing present in this low frequency image[3] that the proposed dictionary cannot efficiently capture.

C. Rate Scalability

MP also offers an intrinsic multiresolution advantage, which can be efficiently exploited for rate adaptivity. The coefficients are by nature exponentially decreasing so that the stream can simply be truncated at any point to provide a SNR-adaptive bitstream, while ensuring that the most energetic atoms are kept. The simplest possible rate adaption algorithm that uses the progressive nature of the MP stream works as follows. Assume an image has been encoded at a high target bit-rate R_b, using the rate controller described in Section 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 adaptation, or filtering operation is equivalent to dropping the last iterations in the MP expansion, focusing on the highest energy atoms.

---

[3]The Daubechies 9/7 filter gives aliasing when reconstructing at a smaller resolution, aliasing which is compensated for the reconstruction of the original size image by information contained in the higher frequency bands.
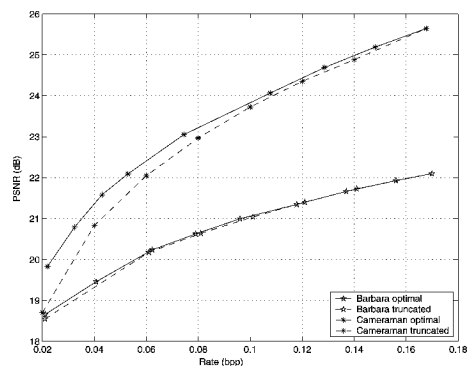

Fig. 12. Rate-distortion characteristics for MP encoding of the 256×256 *Barbara* and *Cameraman* images at 0.17 bpp, and truncation/decoding at different (smaller) bit rates.

Fig. 12 illustrates the rate adaptivity 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. For comparison, the bitstream has also been encoded directly at the different target rates r_k, as described in Section 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 Section III-C. The quantization parameters are not optimal anymore with respect to the truncation rate, but the penalty is quite low away from very low coding rates. The loss in performance 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 adaptation method, though quite basic, is very efficient.

VI. CONCLUSION

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 series of nonseparable, 2-D functions has been emphasized, as a key to overstep limitations of classical representation techniques. The proposed image coder performs a MP 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 is comparable, at low bit rate, with state of the art coders like JPEG2000 or SPIHT, both in terms of PSNR and visual quality. The coding artifacts are in general much less annoying for the MP-based encoder, thanks to the dictionary design.

In addition to good compression performance for low bit-rate coding, the MP coder provides interesting adaptivity features. Due to the structured nature of the dictionary, the bit-stream can be adapted to any spatial resolution at a negligible computational cost. In the same time, the intrinsic multiresolution features of MP expansions allow for very simple rate adaptation. Even if the MP encoder mainly provides good compression performance at low to medium rates, its adaptivity features stay interesting for particular applications, such as huge image databases queried through portable devices, where low bit rate coding is important for quick browsing. They are also interesting in layered coding scenario for example, where the base layer becomes adaptive too. Some work has already been done in order to complement the MP coder with a wavelet coder, in order to achieve very high bit-rate coding [42], while maintaining adaptivity. Extensions of this work to video coding, and particularly efficient frame-rate adaptation methods, are certainly 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 flexible and easily manageable bitstreams. By relaxing these requirements, the coding performance could be largely improved. It can be shown that the proposed representation clearly provides a sparser image approximation than separable decompositions, like wavelet-based schemes. However, this primary advantage is counter-balanced by a very efficient coding strategy in wavelet-based schemes. An improved coding scheme adapted to the proper characteristics of the MP decomposition is currently under study.

REFERENCES

[1] M. N. Do and M. Vetterli, "Contourlets: A directional multiresolution image representation," in *Proc. Int. Conf. Image Processing*, vol. 1, Rochester, NY, Sep. 2002, pp. 357–360.

[2] R. Shukla, P. L. Dragotti, M. Do, and M. Vetterli, "Rate-distortion optimized tree-structured compression algorithms for piecewise polynomial images," *IEEE Trans. Image Process.*, vol. 14, no. 3, pp. 343–359, Mar. 2005.

[3] E. Le Pennec and S. Mallat, "Bandelet representation for image compression," in *Proc. Int. Conf. Image Processing*, vol. 1, 2001, pp. 12–15.

[4] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 13, no. 381(6583), pp. 607–609, Jun. 1996.

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

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

[7] D. L. Donoho, M. Vetterli, R. A. DeVore, and I. Daubechies, "Data compression and harmonic analysis," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 391–432, Mar. 1998.

[8] M. N. Do, P. L. Dragotti, R. Shukla, and M. Vetterli, "On the compression of two dimensional piecewise smooth functions," presented at the Int. Conf. Image Processing, 2001.

[9] R. M. Figueras i Ventura, L. Granai, and P. Vandergheynst, "R-D analysis of adaptive edge representations," in *Proc. MMSP*, 2002, pp. 130–133.

[10] J.-P. Antoine, R. Murenzi, and P. Vandergheynst, "Directional wavelets revisited: Cauchy wavelets and symmetry detection in patterns," *Appl. Comput. Harmon. Anal.*, vol. 6, no. 3, pp. 314–345, 1999.

[11] E. J. Candès and D. L. Donoho, "Curvelets—a surprisingly effective nonadaptive representation for objects with edges," in *Curves and Surface Fitting, Saint-Malo 1999*, A. Cohen, C. Rabut, and L. Schmaker, Eds. Nashville, TN: Vanderbilt Univ. Press., 2000.

[12] J.-L. Starck, E. J. Candès, and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Trans. Image Process.*, vol. 11, no. 5, pp. 670–684, May 2002.

[13] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Computing*, vol. 20, no. 1, pp. 33–61, 1998.

[14] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decompositions," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2845–2862, Nov. 2001.

[15] M. Elad and A. M. Bruckstein, "A generalized uncertainty principle and sparse representations in pairs of bases," *IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 2558–2567, Sep. 2002.

[16] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Trans. Inf. Theory*, vol. 49, no. 6, pp. 3320–3325, Dec. 2003.

[17] S. Mallat and Z. Zhang., "Matching pursuit with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[18] R. A. DeVore and V. N. Temlyakov, "Some remarks on greedy algorithms," *Adv. Comput. Math.*, vol. 5, no. 2–3, pp. 173–187, 1996.

[19] L. K. Jones, "On a conjecture of Huber concerning the convergence of projection pursuit regression," *Ann. Stat.*, vol. 15, no. 2, pp. 880–882, 1987.

[20] L. F. Villemoes, "Nonlinear approximation with Walsh atoms," in *Surface Fitting and Multiresolution Methods*, A. Le Mhaut, C. Rabut, and L. L. Schumaker, Eds. Nashville, TN: Vanderbilt Univ. Press, 1997, pp. 329–336.

[21] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Annu. Asilomar Conf. Signals, Systems, Computers*, vol. 1, ch. , Pacific Grove, CA, Nov. 1993, pp. 40–44.

[22] G. M. Davis, S. G. Mallat, and Z. Zhang, "Adaptive time-frequency decompositions," *SPIE J. Opt. Eng.*, vol. 33, no. 7, pp. 2183–2191, 1994.

[23] A. C. Gilbert, S. Muthukrishnan, and M. J. Strauss, "Approximation of functions over redundant dictionaries using coherence," in *Proc. 14th Annu. ACM-SIAM Symp. Discrete Algorithms*, 2003, pp. 243–252.

[24] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 5, pp. 2231–2242, Oct. 2004.

[25] R. Gribonval and P. Vandergheynst, "On the exponential convergence of matching pursuit in quasi-incoherent dictionaries," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 255–261, Jan. 2006.

[26] F. Bergeaud and S. Mallat, "Matching pursuit of images," in *Proc. Int. Conf. Image Processing*, vol. 1, 1995, pp. 53–56.

[27] R. Neff and A. Zakhor, "Very low bit-rate video coding based on matching pursuits," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 7, no. 1, pp. 158–171, Jan. 1997.

[28] R. M. Figueras i Ventura, P. Frossard, and P. Vandergheynst, "Evolutionary multiresolution matching pursuit and its relations with the human visual system," in *Proc. EUSIPCO*, vol. 2, 2002, pp. 395–398.

[29] R. M. Figueras i Ventura, O. D. Escoda, and P. Vandergheynst, "Matching Pursuit Full Search Algorithm," ITS (EPFL), Tech. Rep. TR-ITS-31/2004, 2004.

[30] D. Marr, *Vision*. San Francisco, CA: Freeman, 1982.

[31] D. Bernier and K. Taylor, "Wavelets from square-integrable representations," *SIAM J. Math. Anal.*, vol. 27, no. 2, pp. 594–608, 1996.

[32] P. Frossard, P. Vandergheynst, R. M. Figueras i Ventura, and M. Kunt, "A posteriori quantization of progressive matching pursuit streams," *IEEE Trans. Signal Process.*, vol. 52, no. 4, pp. 525–535, Apr. 2004.

[33] S. Mallat, *A Wavelet Tour of Signal Processing*. New York: Academic, 1998.

[34] I. H. Witten, R. M. Neal, and J. G. Cleary, "Arithmetic coding for data compression," *Commun. ACM*, vol. 30, no. 6, pp. 520–540, 1987.

[35] D. L. Duttweiler and C. Chamzas, "Probability estimation in arithmetic and adaptive-Huffman entropy coders," *IEEE Trans. Image Process.*, vol. 4, no. 2, pp. 237–246, Feb. 1995.

[36] Z. Cvetkovic and M. Vetterli, "Tight Weyl-Heisenberg frames in $l^2(z)$," *IEEE Trans. Signal Process.*, vol. 46, no. 9, pp. 1256–1259, Sep. 1998.

[37] A. Said and W. Pearlman, "Reversible image compression via multiresolution representation and predictive coding," in *Proc. SPIE VCIP*, 1993, pp. 664–674.

[38] P. Vandergheynst and P. Frossard, "Efficient image representation by anisotropic refinement in matching pursuit," in *Proc. ICASSP*, vol. 3, 2001, pp. 1757–1760.

[39] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 6, no. 2, pp. 243–250, Feb. 1996.

[40] D. Taubman and A. Zakhor, "Multirate 3-d subband coding of video," *IEEE Trans. Image Process.*, vol. 3, no. 4, pp. 572–588, Apr. 1994.

[41] J. W. Woods and G. Lilienfield, "A resolution and frame-rate scalable subband/wavelet video coder," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 8, pp. 1035–1044, Aug. 2001.

[42] L. Peotta, L. Granai, and P. Vandergheynst, "Very low bit rate image coding using redundant dictionaries," in *Proc. SPIE, Wavelets: Applications Signal Image Processing X*, vol. 5207, 2003, pp. 228–229.

Pierre Vandergheynst (M'01) received the M.S. degree in physics and the Ph.D. degree in mathematical physics from the Université Catholique de Louvain, Louvain, Belgium, in 1995 and 1998, respectively.

From 1998 to 2001, he was a Postdoctoral Researcher with the Signal Processing Laboratory, Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland. He is now an Assistant Professor of visual information representation theory at EPFL, in charge of the LTS2 Laboratory, where his research focuses on active in several research topics ranging from time-frequency/time-scale analysis to mathematical image analysis, computer vision, data compression, and higher dimensional data processing.

Dr. Vandergheynst is co-Editor-in-Chief of *Signal Processing*.

Pascal Frossard (S'96–M'01–SM'04) received the M.S. and Ph.D. degrees, both in electrical engineering, from the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, in 1997 and 2000, respectively. In 1998, he graduated from the Graduate School of the Communications System Division of the EPFL.

From 1998 to 2000, he worked with the Signal Processing Laboratory, EPFL, as a research and teaching assistant under a grant from Hewlett-Packard. Between 2001 and 2003, he was a member of the research staff at the IBM T. J. Watson Research Center, Yorktown Heights, NY, where he worked on media compression and streaming technologies. Since April 2003, he has been an assistant professor at EPFL. His research interests include image representation and coding, nonlinear representations, visual information analysis, joint source and channel coding, multimedia communications, and multimedia content distribution. He works on nonlinear signal transforms applied to visual information, and adaptive media streaming algorithms. In 2002, he was co-editor of a Special Issue on new visual coding methods for *Signal Processing*.

Dr. Frossard has been the General Chair of IEEE ICME 2002 (Lausanne, Switzerland), and member of the organizing or technical program committees of numerous conferences. He has served as Guest Editor of special issues on Streaming Media (IEEE TRANSACTIONS ON MULTIMEDIA), on Media and Communication Applications on General Purpose Processors: Hardware and Software Issues *(Journal of VLSI SPSS)*, and on Image and Video Coding Beyond Standards *(Journal of Signal Processing)*. He is an Associate Editor of the IEEE TRANSACTIONS ON MULTIMEDIA (2004–) and of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY (2006–), and he served as a member of the Editorial Board of the *EURASIP Journal of Signal Processing* (2003–2005). Since 2004, he has served as Vice-Chair of the IEEE Multimedia Communications Technical Committee, as a member of the IEEE Multimedia Signal Processing Technical Committee, and of the IEEE Multimedia Systems and Applications Technical Committee. He received the Swiss NSF Professorship Award in 2003, and the IBM Faculty Award in 2005

Rosa M. Figueras i Ventura received the M.S. degree in telecommunication engineering from the Escola Tècnica Superior d'Enginyers de Telecomunicació de Barcelona, Technical University of Catalonia, Barcelona, Spain, in 2000. She is currently pursuing the Ph.D. degree at the Signal Processing Institute, Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland.

During the 1999 to 2000 academic year, she was an exchange student at the EPFL, where she did her M.S. thesis on image coding. While studying for the Ph.D. degree, she has been a Teaching and Research Assistant. Her main current research interests are image processing and representation and the study of the properties of sparse signal approximations.