High flexibility scalable image coding

Pascal Frossard, Pierre Vandergheynst and Rosa M. Figueras i Ventura
Signal Processing Institute
Swiss Federal Institute of Technology, Lausanne, Switzerland

ABSTRACT

This paper presents a new, highly flexible, scalable image coder based on a Matching Pursuit expansion. The dictionary of atoms is built by translation, rotation and anisotropic refinement of gaussian functions, in order to efficiently capture edges in natural images. In the same time, the dictionary is invariant under isotropic scaling, which interestingly leads to very simple spatial resizing operations. It is shown that the proposed scheme compares to state-of-the-art coders when the compressed image is transcoded to a lower (octave-based) spatial resolution. In contrary to common compression formats, our bit-stream can moreover easily and efficiently be decoded at any spatial resolution, even with irrational re-scaling factors. In the same time, the Matching Pursuit algorithm provides an intrinsically progressive stream. This worthy feature allows for easy rate filtering operations, where the least important atoms are simply discarded to fit restrictive bandwidth constraints. Our scheme is finally shown to favorably compare to state-of-the-art progressive coders for moderate to quite important rate reductions.

Keywords: Scalability, progressive coding, transcoding, hierarchical coder, image compression

1. INTRODUCTION

Adaptivity and on-the-fly resolution switching is becoming an important requirement in many visual applications involving scalable transmission and storage, like database browsing or image and video communications. 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 spatial 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.[1,2] 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[3] 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. This paper presents a Matching Pursuit image encoder based on a dictionary of anisotropic and oriented atoms, and emphasizes the intrinsic spatial and rate scalability of the created bitstreams. First, due to the structure of the proposed dictionary, the stream can easily and efficiently be decoded at any spatial resolution, and by extension to any spatio-temporal resolution in the video case. 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 important components of the image.

Pascal Frossard was with the IBM TJ Watson Research Center, Yorktown Heights, NY, USA.
Further author information: (Send correspondence to Pascal Frossard)
Pascal Frossard: E-mail: pascal.frossard@epfl.ch
Pierre Vandergheynst: E-mail: pierre.vandergheynst@epfl.ch
Rosa Figueras i Ventura: E-mail: rosa.figueras@epfl.ch

This paper is organized as follows. Section 2 briefly describes the new Matching Pursuit based image coder used throughout the paper. Section 3 then presents the spatial scalability feature of the proposed coder, while Section 4 discusses the rate adaptivity offered by the Matching Pursuit algorithm. Some applications of the highly flexible adaptivity features of the proposed encoding scheme are listed in Section 5. Section 6 then concludes the paper.

2. MATCHING PURSUIT IMAGE CODER

2.1. Overview

The study and design of a new image coder result from the need for both an efficient (very) low bit-rate image representation and flexible progressive bit-streams. Redundant multidimensional expansions surely represent the core of new breakthroughs in image compression. The optimal decomposition of an image over an overcomplete dictionary is however an NP-complex problem, except in the recently investigated case of *incoherent* dictionaries where it has been shown that Basis Pursuit would actually find the optimal sparse solution.[4, 5] Several algorithms can be proposed to compute the image representation in a sparse set of multidimensional functions, and the choice of a particular algorithm generally consists in trading off complexity and efficiency. The new image compression scheme presented in this paper proposes to use Matching Pursuit as an efficient way to produce a progressive low bit-rate image representation with a controlled complexity.

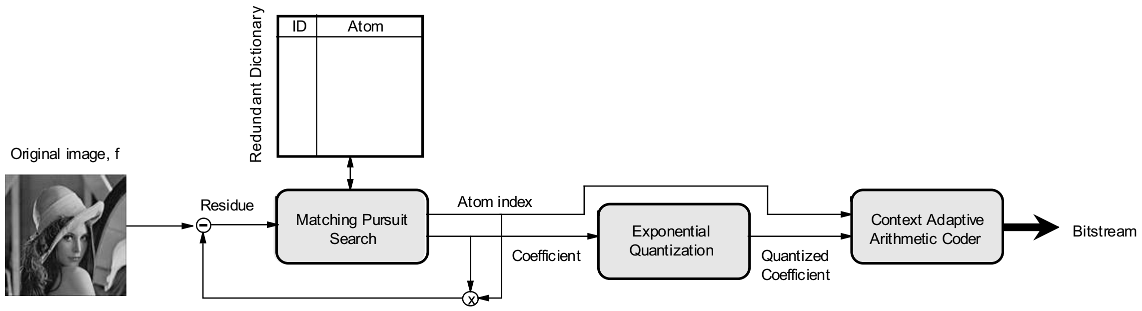

Figure 1. Block diagram of the Matching Pursuit image coder.

The compression scheme can be represented as in Figure 1. 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.[6] Coefficients and atom indexes are finally entropy coded with a context adaptive arithmetic coder. The image is reconstructed by performing the reverse operations at the decoder, and is successively refined by each additional bits of information. One of the interesting characteristics of Matching Pursuit expansions resides in their ability to provide meaningful image representations even with very short descriptions of the input signal.

2.2. Anisotropic refinement using Matching Pursuit

Matching Pursuit can be defined as *a greedy algorithm that decomposes any signal into a linear expansion of waveforms taken from a redundant dictionary*.[7] These waveforms are iteratively chosen to best match the signal structures, producing a sub-optimal non-linear signal decomposition. The N-term Matching Pursuit expansion of the signal f can be written as :

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

where the residual of the decomposition $\|R^N f\|$ converges exponentially to 0 when N tends to infinity, for a finite dimensional signal f. The greedy Matching Pursuit decomposition moreover conserves the L^2 norm of the signal f in the transformed domain, which yields :

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

Matching Pursuit coding efficiency is however highly dependent on the dictionary, and its ability to capture the characteristics of the input signal f. The dictionary proposed in this paper is built by applying meaningful geometric transformations to a generating function of unit L^2 norm. These transformations can be represented by a family of unitary operators U_γ, and the dictionary is thus expressed as :

$$\mathcal{D} = \{U_\gamma,\ \gamma \in \Gamma\}\ , \tag{3}$$

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.
- Anistropic 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 :

$$U_\gamma g = \mathcal{U}(\vec{b}, \theta) D(a_1, a_2) g \tag{4}$$

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

$$\mathcal{U}(\vec{b}, \theta) g(\vec{x}) = g\big(r_{-\theta}(\vec{x} - \vec{b})\big)\ , \tag{5}$$

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

$$D(a_1, a_2) g(x, y) = \frac{1}{\sqrt{a_1 a_2}} g\big(\frac{x}{a_1}, \frac{y}{a_2}\big)\ . \tag{6}$$

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.[8] It is also worth stressing that, avoiding rotations, the parameter space is a group studied by Bernier and Taylor.[9] 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. $a_1 = a_2$; this property will be exploited for spatial transcoding in the next sections. The invariance with respect to dilations is useful when expanding a small image, providing quite good performance even for irrational up-scaling.

The choice of the generating atom g is driven by the idea of efficiently approximating contour-like singularities in 2-D. To achieve this goal, the atom must be a smooth low resolution function in the direction of the contour and must behave like a wavelet in the orthogonal (singular) direction. Our atoms are built on gaussian functions along the first direction and second derivative of gaussian functions in the orthogonal direction. This choice is motivated by the optimal joint spatial and frequency localization of the gaussian kernel. We also noticed that degradation caused by truncating the Matching Pursuit expansion are visually less disturbing with this choice of function.

Finally, the dictionary used in this paper is called *structured* dictionary, in the sense that it 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 :

$$g_\gamma = (4\ {g_1}^2 - 2)\ \exp(-({g_1}^2 + {g_2}^2))\ , \tag{7}$$

with

$$g_1 = \frac{\cos(\theta)\ (x - b_1) + \sin(\theta)\ (y - b_2)}{a_1}\ , \tag{8}$$

and

$$g_2 = \frac{\cos(\theta)\ (y - b_2) - \sin(\theta)\ (x - b_1)}{a_2}\ . \tag{9}$$

3. SPATIAL SCALABILITY

Thanks 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 and allows for covariance 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 :

$$\mathcal{U}(\eta)\hat{f} = \sum_{n=0}^{N-1} \langle g_{\gamma_n} | \mathcal{R}^n f \rangle \, g_{\eta \circ \gamma_n}\ . \tag{10}$$

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 denote 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 :

$$\mathcal{U}(\eta_\alpha)\hat{f} = \alpha\ \sum_{n=0}^{N-1} c_{\gamma_n} g_{\eta_\alpha \circ \gamma_n}\ . \tag{11}$$

The transcoded atoms $g_{\eta_\alpha \circ \gamma_n}$ are given by Eqs. (7) to (9), where $\vec{b}$ and $\vec{a}$ are respectively replaced by $\alpha\ \vec{b}$ and $\alpha\ \vec{a}$. Atoms that becomes too small after transcoding can be discarded, allowing for further bit-rate reduction. It is worth noting that the scaling factor α can take any real value, as long as the scaling is isotropic. Finally, image editing manipulations, such as rotation of the image, or zoom in a region of interest, can easily be implemented following the same principle.

Figure 2. *lena* image of size 128×128 encoded with MP at 1.6bpp (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. 2, where the encoded image of size 128×128 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 in the compressed bit-stream.

Encoder	32x32	64x64	128x128	256x256	512x512
Matching Pursuit [dB/bpp]	16.7/4	18.97/2.25	30.37/1.7	26.05/0.41	25.94/0.1
JPEG-2000 [dB/bpp]	16.98/6.5	19.18/4	33.8/1.7	-	-

Table 1. Comparison of spatial scalability of the MP encoder and JPEG2000.

Table 3 shows rate-distortion performance of spatial resizing of the 128×128 *lena* image compressed with the proposed Matching Pursuit coder, and JPEG2000 respectively. In addition to allowing for non-dyadic spatial resolutions, as well as easy up-scaling, our scheme offers quite competitive results with respect to state-of-the-art coders like JPEG2000 for octave-based resizing.

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. In these experiments, the PSNR values have been computed with reference to the original 512×512 pixels *lena* image, downsampled to 128×128 pixels. This is one possibility for computing such a reference and other more complex techniques, involving for example filtering and interpolation, could be adopted. They would however not significantly change the comparative results.

4. RATE SCALABILITY

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 to provide a SNR-scalable bitstream, while ensuring that the most energetic atoms are saved.

Figure 3. *lena* image of size 128×128 encoded with MP at 1.7bpp (top left), and truncated at 0.1bpp (top right), 0.4 bpp (bottom left), and 0.8 bpp (bottom right).

In this paper we propose the simplest possible rate transcoding algorithm that uses the progressive nature of the Matching Pursuit stream. Assume an image has been encoded at a high target bit-rate R, using the rate controller described in.[6] 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. 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 4. *lena* image of size 128 × 128 encoded with JPEG-2000 at 1.7 bpp (top left), and truncated at 0.1bpp (top right), 0.4 bpp (bottom left), and 0.8 bpp (bottom right).

Decoded images based on this procedure are provided in Figure 3. The *lena* has been encoded at 1.7 bpp, and decoded at respectively 0.1, 0.4 and 0.8 bpp by truncation of the bit-stream. For the sake of comparison, Figure 4 shows the result of the same filtering operation on a progressive JPEG2000 stream. Visual inspection of these images shows that our scheme provides smoother approximations and visually meaningful details, even for very low rates. In particular there is no annoying ringing artifacts on the MP streams. The low rate image is very sketchy and more details are being added as the bit budget r_k is increased. This is easily explained by the fact that the very first few atoms in the MP stream carry most of the energy and most of the visual information.

Figure 5 finally shows a rate-distortion comparison of our technique with respect to the JPEG2000 standard. The *lena* image has been encoded with MP at a high rate of 1.7 bpp and truncated to lower rates r_k. The same experiments have been performed with JPEG2000, and the bitstream has been decoded at the same target rates r_k. The image has also been directly encoded with MP at the successive target rates r_k. As one can see on these curves, the truncated streams are always better than the JPEG2000 ones at low bit rates. Of course, there is always a loss in PSNR with respect to the optimal MP stream at the same rate, since the truncation simply results in dropping iterations, without using the optimal quantizer settings imposed by rates r_k as proposed in.[6] Nevertheless, both optimal and truncated rate-distortion curves are quite close, which shows that our simple rate transcoding method, though basic, is very efficient.

5. APPLICATIONS

There are obviously numerous potential applications of this technique and we will just name a few here.

First of all, the spatial scalability of our coder can be used to speed-up computations, an endemic problem of non-linear image representations such as Matching Pursuit. Indeed, as explained in,[10] one can start by encoding

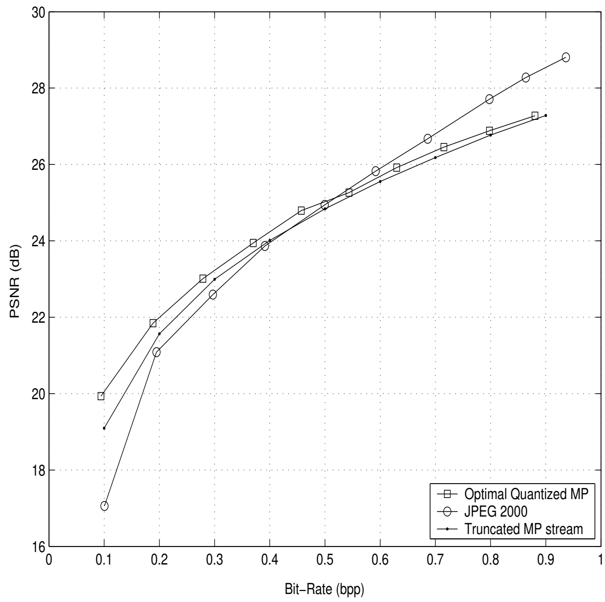

Figure 5. Rate-distortion characteristics for MP and JPEG2000 encodings at 1.7 bpp, and truncation/decoding at different (smaller) bit rates.

a downsampled version of the image using Matching Pursuit, and then directly scale this approximation up to a higher resolution using Eq. (11). A new residual is then computed by subtracting this reconstructed, up-scaled, image from the original image downsampled at the corresponding resolution. The process is then iterated until we reach the full image size. The main advantage of this technique is that most of the energy is generally concentrated at low resolutions and this is encoded by the first call to Matching Pursuit. Since the algorithm is applied to a small downsampled image, the number of computations is much lower than for a Matching Pursuit expansion of the full resolution image. A few iterations only are necessary at subsequent finer resolutions in order to encode the remaining high frequency characteristics.

Another class of interesting applications arises in the context of image database exploration, where fast access to a low resolution approximations (i.e. thumbnail version) of the stored images allows for swift browsing of the whole collection. In this problem, good visual quality at low bit rates is mandatory and it is precisely one of the main advantages of our technique. Multiple resolutions can be generated from a single bit-stream, thus avoiding costly storage of numerous compressed bit-stream of the same image.

When dealing with visual communications, a common networking problem is the heterogeneity of both receivers and channel bandwidths. Our scheme has the advantage of being able to solve both problems at once by efficiently adapting to any spatial resolution and any rate constraint. Moreover, transcoding operations are kept very simple and do not require heavy post-processing or bitstream manipulations in the network nodes. We are thus in a position to embed significant network intelligence at a reduced computational cost.

Finally, our results are based either on inherent Matching Pursuit features or on geometric properties of the dictionary. They can thus readily be extended to spatio-temporal dictionaries and used for video coding. Indeed, by building a dictionary based on meaningful spatio-temporal operations (i.e. spatial operations extended by temporal translation and scaling), we obtain efficient scalable moving pictures representations. In these settings, spatial, rate and also temporal scalability, can be tailored into progressive streams using the arguments developed in this paper. The interested reader may consult[11] for more details.

6. CONCLUSIONS

We have presented a high flexibility adaptive image coding scheme based on a new Matching Pursuit image encoder. Thanks to the structured nature of the dictionary, the bit-stream can be transcoded to any spatial resolution at a very low computational cost. In the same time, the intrinsic multi-resolution feature of Matching Pursuit expansions allows for very simple rate adaptive operations.

Our scheme has been shown to provide performance similar to JPEG2000 for dyadic spatial resolutions. It however offers much more flexibility and even allows for irrational re-scaling factors. In the same time, the Matching Pursuit based scheme outperforms state-of-the-art compression schemes in the case of (very) low bit rate transcoding.

The very nice scalability features of multidimensional expansions such as the one proposed in this paper open interesting perspectives for numerous visual communication applications, where the heterogeneity of the receivers is an important problem. Extensions of this work to video coding, and particularly efficient frame-rate adaptivity methods, are worth investigating as a potential solution to adaptive video delivery scenarios.

REFERENCES

1. 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* **6**, pp. 243–250, June 1996.
2. Taubman D. and Zakhor A., "Multirate 3-D Subband Coding of Video," *IEEE Transactions on Image Processing* **3**, pp. 572–588, September 1994.
3. 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* **11**, pp. 1035–1044, September 2001.
4. Elad M. and Bruckstein A.M., "A generalized uncertainty principle and sparse representations in pairs of bases," *IEEE Transactions on Information Theory* **48**, pp. 2558–2567, September 2002.
5. Donoho D.L. and Huo X., "Uncertainty principles and ideal atomic decompositions," *IEEE Transactions on Information Theory* **47**, pp. 2845–2862, November 2001.
6. 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* , February 2003. accepted for publication.
7. Mallat S.G. and Zhang Z., "Matching Pursuits With Time-Frequency Dictionaries," *IEEE Transactions on Signal Processing* **41**, pp. 3397–3415, December 1993.
8. Antoine J.-P., Murenzi R. and Vandergheynst P., "Directional wavelets revisited: Cauchy wavelets and symmetry detection in patterns," *Applied and Computational Harmonic Analysis* **6**, pp. 314–345, 1999.
9. Bernier D. and Taylor K., "Wavelets from square-integrable representations," *SIAM Journal on Mathematical Analysis* **27**(2), pp. 594–608, 1996.
10. 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*, **2**, pp. 395–398, (Toulouse, France), September 2002.
11. Frossard P., *Robust and Multiresolution Video Delivery: From H.26x to Matching Pursuit Based Technologies.* PhD thesis, Swiss Federal Institute of Technology, Lausanne, Switzerland, December 2000.