EVOLUTIONARY MULTIRESOLUTION MATCHING PURSUIT AND ITS RELATIONS WITH THE HUMAN VISUAL SYSTEM

Rosa M. Figueras i Ventura[1], *Pierre Vandergheynst*[1] *and Pascal Frossard*[2]
[1] Signal Processing Laboratory,
Swiss Federal Institute of Technology (EPFL),
CH-1015 Lausanne,
Switzerland
{rosa.figueras, pierre.vandergheynst}@epfl.ch

[2] IBM TJ Watson Research Center
30, Saw Mil River Road
Hawthorne, NY 10532,
USA
frossard@us.ibm.com

ABSTRACT

This paper proposes a multiresolution Matching Pursuit decomposition of natural images. Matching Pursuit is a greedy algorithm that decomposes any signal into a linear expansion of waveforms taken from a redundant dictionary, by iteratively picking the waveform that best matches the input signal. Since the computational cost rapidly grows with the size of the signal, we propose a multiresolution strategy that, together with a dictionary training, significantly reduces the encoding complexity while still providing an efficient representation. Such a decomposition is perceptually very effective at low bit rate coding, thanks to similiarities with the Human Visual System information processing.

1 INTRODUCTION

Matching Pursuit use is spreading in image and video coding due to its good profile in very low bit-rate applications [2, 4] and in denoising [1]. MP good performance in very low bit-rate applications comes mainly from two factors: Its non-linearity, which allows to better represent a signal with a lower number of terms, and its similarity with the Human Visual System, which will be pointed out further in this paper. MP non-linearity allows for detecting the main contours of an image with a very low number of terms.

2 MATCHING PURSUIT

2.1 The algorithm

The basis of Matching Pursuit can be found in Mallat [9] and Mallat and Zhang [10]). They define Matching Pursuit as *a greedy algorithm that decomposes any signal into a linear expansion of waveforms taken from a redundant dictionary*. These waveforms are iteratively chosen to best match the signal structures, producing a sub-optimal expansion. Vectors are selected one by one from the dictionary, while optimizing the signal approximation (in terms of energy) at each step. Even though the expansion is linear, it gives a non-linear signal decomposition.

Let $\mathcal{D} = \{g_\gamma\}_{\gamma\in\Gamma}$ be a dictionary of $P > M_1 \times M_2$ vectors, having unit norm. This dictionary includes $M_1 \times M_2$ linearly independent vectors that define a basis of the space $\mathbb{C}^{M_1\times M_2}$ of signals of size $M_1 \times M_2$. Let $R^n f$ be the residual of an n term representation of a given signal f.

A Matching Pursuit is an iterative algorithm that subdecomposes the residue $R^n f$ by projecting it on a vector of $\mathcal{D}$ that matches $R^n f$ at best. If we consider $R^0 f = f$ first MP iteration will represent the signal as:

$$f = R^0 f = \langle f, g_{\gamma_0}\rangle g_{\gamma_0} + R^1 f, \tag{1}$$

where $R^1 f$ is the residual vector after approximating $R^0 f$ in the direction of g_{γ_0}. Since $R^1 f$ is orthogonal to g_{γ_0}, the module of f will be:

$$\|R^0 f\|^2 = |\langle R^0 f, g_{\gamma_0}\rangle|^2 + \|R^1 f\|^2. \tag{2}$$

As the term that must be minimized is the error $\|R^1 f\|$,

$$\|R^1 f\|^2 = \|R^0 f\|^2 - |\langle R^0 f, g_{\gamma_0}\rangle|^2, \tag{3}$$

the $g_\gamma \in \mathcal{D}$ to be chosen is the one that maximizes $|\langle R^0 f, g_{\gamma_0}\rangle|$, or, generalizing, $|\langle R^n f, g_{\gamma_n}\rangle|$. In some cases it is not computationally efficient to find the optimal solution, and a suboptimal solution is computed instead:

$$|\langle R^n f, g_{\gamma_0}\rangle| \geqslant \alpha \sup_{\gamma\in\Gamma} |\langle R^n f, g_\gamma\rangle|, \tag{4}$$

where $\alpha \in (0, 1]$ is an optimality factor which is 1 when the optimal solution has been chosen. This suboptimality factor α will depend on the searching method used to find the solution (see section 3 for an example).

From (1), one easily sees by induction that the N term decomposition of f is given by:

$$f = \sum_{n=0}^{N-1} \langle R^n f, g_{\gamma_n}\rangle g_{\gamma_n} + R^N f \tag{5}$$

and with the same principle we can also deduce from (2) that the L^2 norm of the signal f is:

$$\|f\|^2 = \sum_{n=0}^{N-1} |\langle R^n f, g_{\gamma_n}\rangle|^2 + \|R^N f\|^2, \tag{6}$$

where $\|R^N f\|$, when dealing with finite dimension signals, converges exponentially to 0 when N tends to infinity and $M_1 \times M_2$ is finite (see [7] for a proof).

Matching Pursuit coding efficiency is highly dependent on the dictionary adaptation to the signal to represent. In the next section we describe a general framework for handling geometric dictionary construction.

Figure 1: Anisotropic refinement atoms versus Isotropic Gabor atoms. Anisotropy (right image) gives better contour resolution.

2.2 Properties of Matching Pursuit

MP properties can be divided in two different kinds: properties that are intrinsical to the algorithm, no matter which kind of functions have been used to perform the signal decomposition, and properties that depend of the dictionary (they appear only if the dictionary used has them as well).

The main properties derived directly from the Matching Pursuit algorithm are *invertibility* (if the dictionary is at least complete), *energy conservation* (that comes from Eq. 6 and the invertibility property) and *overcompleteness*, which gives *robustness to quantization* (due to the fact that the coding space is of higher dimension than the signal space) and *exponentially bounded error decay* (which implies a fast initial error decay).

Most of the other properties depend on the dictionary at hand. In particular covariance with respect to geometric transformations is a very desirable feature. Let us quickly explain a generic way of achieving such a construction.

Suppose we have a group of geometric transformations G together with a unitary representation $\mathcal{U}$ of G in the Hilbert space of our signals $\mathcal{H}$. It is a classical result of group representation theory [3] that the dictionary

$$\mathcal{D} = \{\mathcal{U}(\gamma)g\,,\ \forall\gamma \in G\}$$

is a dense subspace of $\mathcal{H}$ for any $g \in \mathcal{H}$. This dictionary is invariant under any geometric transformation in G by construction. Moreover the MP expansion of any deformed signal is very simply related to the expansion of the original signal :

$$\mathcal{U}(\gamma)f = \sum_{n=0}^{+\infty} \langle g_{\gamma_n} | R^n f \rangle g_{\gamma\circ\gamma_n}\ ,$$

where $\circ$ denotes group composition. Summarizing, we have complete invariance of the MP expansion with respect to geometric transformations. Group transformations can be combined with more general manipulation for creating special dictionaries. Suppose we create a dictionary by applying both a unitary group representation $\mathcal{U}$ and another unitary operator R_α to a generating function g :

$$\mathcal{D} = \{\mathcal{U}(\gamma)\cdot R_\alpha g\, \forall\gamma \in G\ \alpha \in A\}\ . \qquad (7)$$

Then this dictionary would still benefit from the invariance properties of G. A very fruitful example used throughout the remaining of this paper is obtained by taking G as composed of translations and rotations and extended as in Eq. (7) with anisotropic dilations :

$$R(a_1,a_2)g(x,y) = \frac{1}{\sqrt{a_1 a_2}} g\left(\frac{x}{a_1}, \frac{y}{a_2}\right)\ .$$

This dictionary is then invariant under translations, rotations and *isotropic* scaling $a_1 = a_2$. In our case the dictionary used is composed of Anisotropic Refinement atoms. The basic function is a Gaussian in one axis and the second derivative of the Gaussian in the other axis [13]:

$$g_\gamma(x,y) = (2-4x^2)e^{\frac{-1}{4}(x^2+y^2)}. \qquad (8)$$

This particular set of atoms is very well suited for representing smooth contours in images as already pointed out in [13] and this fact is also illustrated on Figure 1.

3 EVOLUTIONARY MATCHING PURSUIT

The use of a redundant basis through MP seems interesting from an image representation point of view, but it represents a heavy computational cost. In fact, when dealing with large dictionaries, the computation of a scalar product of every element of the dictionary and the signal to represent and take the atom with the largest projection energy becomes almost impossible. In this scope, the use of efficient approximation tools, such as Genetic Algorithms, is needed.

GA do not give the optimal solution, but an approximation. This fact, though, does not represent a problem when dealing with MP decomposition. It will, of course, cause a decrease of quality in the final MP representation, but this loss of quality is negligible compared to the computational gain obtained.

The GA used here is a simple algorithm that has demonstrated to be well adapted to the needs of this concrete case. It has a population formed by a certain odd number N of individuals. Each individual is in fact one dictionary component, and it is composed by five genes (which are the parameters that define the dictionary component, so, position in x and y, scaling in x and y and rotation). At every generation, these individuals are evaluated and only the fittest (the one that has higher scalar product norm) passes to the next generation without change. The rest compete in pairs, and the winner of every pair is placed in a matching pool. The individuals in the matching pool are randomly crossed-over, and their descendants ($\frac{N-1}{2}$ individuals) are placed at the next generation together with $\frac{N-1}{2}$ mutations of the fittest. The evaluation process is repeated until a desired error threshold or a certain number of generations has been reached (see [8] for a detailed description of GA).

4 MULTIRESOLUTION MATCHING PURSUIT

As the search for the optimal function means computing a great amount of scalar products between images, MP has a very high computational cost, which directly depends on the image size. To speed coding, a multiresolution scheme has

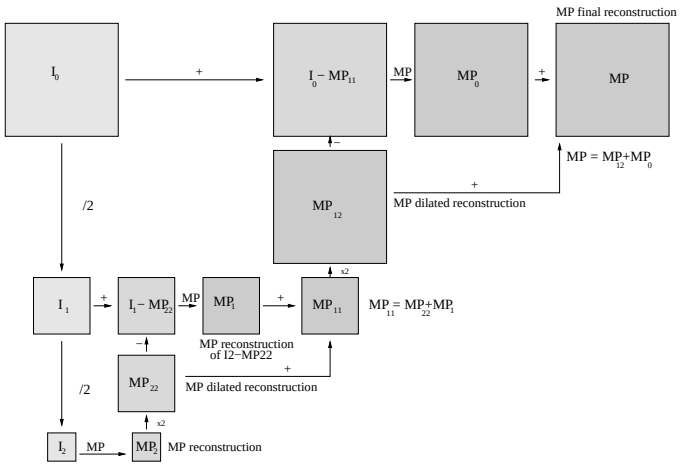

Figure 2: Multiresolution MP scheme.

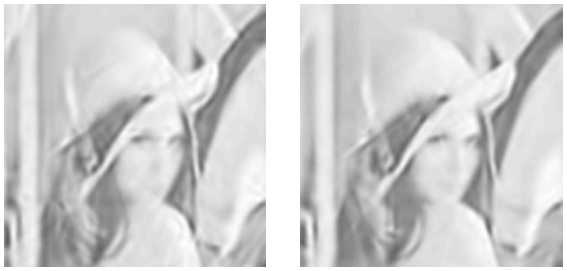

Figure 3: Comparison of two MP coded images. Left image has been coded with one layer MP (PSNR=25.7987 dB) and the right one with multiresolution MP (PSNR=26.0016 dB).

been chosen. In this scheme, the image is downsampled by two several times. The MP algorithm is first applied to the smallest image and when the desired number of coefficients in the lowest resolution layer has been reached, a recomposition of the next level image (double size) is performed. This recomposition is done by taking advantage of the dictionary covariance to dilations (see section 2.2). The subtraction of this recomposition to the next resolution level image is performed, and MP is applied to this residual (see scheme in Fig. 2).

Multiresolution MP normally gives better results that one resolution MP (see Fig. 3), basically because multiresolution decreases the number of atoms in the dictionary for the searching algorithm (the scaling factor is now bounded to the smallest image size). With the same number of generations in the GA, the obtained solution is thus closer to the optimal. A clear study of GA parameters at each level of resolution has got to be performed. The optimal number of terms in the MP expansion at each resolution is also currently investigated.

5 MP IN A LEARNT DICTIONARY SET

As the MP dictionary used here is highly redundant, some of the functions are hardly ever used, they can be rejected (as done by Neff and Zakhor in [11]). The rejection of these

Figure 4: Comparison of two images having 420 coefficients. Left image is coded with the whole dictionary (PSNR=31.2294 dB), and the right one with a subset of 64 functions, learnt from MP decomposition of ten images (PSNR=29.2315 dB).

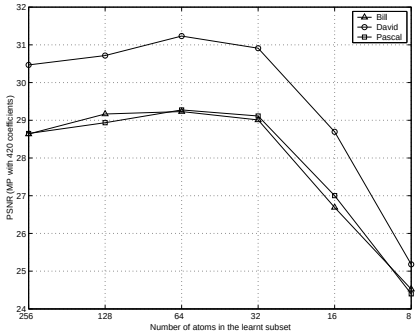

Figure 5: Graphic of the PSNR evolution with the diminution of the size of the subset.

functions will give a more compact representation and reduce the bit-rate to code an image. One way of doing it is decomposing a representative number of images in the whole dictionary, and then train the dictionary from the decomposition of these images. Then, MP can be applied in a learnt subset of functions. This gives higher compression rates (an atom is just represented by an index in a list). It will also speed the Matching Pursuit process because there will be less elements to compare.

A possible learning rule for the subset could be to choose the atoms that appear more often. So, if a subdictionary of M atoms is desired, the first M atoms that most often appear in a representative set of MP decompositions would be chosen. This, though, gives an incorrect result, because some atoms that have a very small impact in the final result are chosen. The reason is that the atoms that appear more often are usually the ones that come after a high number of iterations and the energy they bring to the final result is very small.

To avoid this, a learning rule based on the energy the atom gives to the final reconstruction has been used. The new learning rule will select the first M atoms that have higher energy. One example of MP decomposition using a 64 atoms subset is shown in figure 4.

The quality of the image coded in a subset will depend on the number of atoms this subset has. Intuitively, image quality will increase with the size of the subdictionary used. This is true when using an algorithm which finds the optimal so-

lution, but the use of a suboptimal minimization algorithm changes this logical evolution. Decreasing the number of functions in the chosen subset means reducing the searching space. Logically, when reducing the searching space, with the same complexity in the approximation algorithm, the solution found will be closer to the real one. With the GA used in the scope of this paper, and 128x128 images, the best compromise subset size versus algorithm accuracy has proved to be when taking a 64 atoms subset (see Fig. 5).

6 SIMILARITIES OF MRMP AND HVS

Multiresolution Matching Pursuit (MRMP) shares some interesting properties with the Human Visual System (HVS). These similarities explain, to some extent, why MRMP coded images often have a better visual quality than their wavelet or DCT equivalent even for lower PSNR values.

MRMP similarities with the HVS come in two main flavors : those directly coming from the algorithm and those depending on the particular dictionary used.

One of the first goals of the HVS is to perform a sparse coding of visual information [12]. By nature MRMP yields a very sparse coding of images since, as already pointed before, PSNR increases quickly with the number of coefficients used in the expansion. In fact MRMP seeks particular structures in the image and will recursively extract them from the data. At very low bit rates, or for few coefficients, the selected atoms tend to be independent from each other. This gain in information will of course saturate as the number of terms gets bigger, but quickly yields a good and very sparse approximation of the data. In this way MRMP is more oriented towards meaningful structures where wavelets and DCT merely see pixels.

Concerning the dictionary, several links with the HVS can be highlighted. First we know that visual information is submitted to a chain of processing. At an early stage retinal ganglion cells detect contours using a strategy that mimics the zero-crossings of a Laplacian pyramid. At later stages the information is processed in the primary visual visual cortex (area V1) by several neural cells. Among these, Simple Cells have a receptive field that has been shown to be well approximated by Gabor filters [6]. They are sensitive to the position, scale and orientation of stimuli. Now in the MRMP algorithm presented in section 4, the image is first decomposed at low resolution and this coarse approximation is subtracted from a finer level of the pyramid. This result in a scheme similar to the Laplacian pyramid of Burt and Adelson [5] and thus mimics the early processing stage of the HVS. Then at each resolution, MRMP uses a dictionary of atoms that are sensitive to the position, scale and local orientation of contours, which again resembles some of the processing achieved by Simple Cells. Finally the anisotropic scaling of our dictionary allows us to represent contour information with few atoms by locally stretching the atom in the direction of the edge, while we refine it in the direction of the gradient.

7 CONCLUSIONS

We introduced MRMP, an algorithm that uses Matching Pursuit in a multiresolution fashion with a dedicated dictionary of scale-covariant atoms. Sparse coding of natural images is achieved within our algorithm by using a mixture of properties that mimic sparse structure coding in the Human Visual System. This allows to obtain superior visual quality when compared to traditional linear transforms such as wavelet or DCT.

Such a technique, though at a preliminary stage, might yield very interesting results when combined with efficient coding strategies for very low bit rate image compression.

References

[1] Bruno Aiazzi and Luciano Alparone. Blind image estimation through fuzzy matching pursuits. In *ICIP, IEEE Proceedings*, Thessaloniki, Greece, October 2001.

[2] Osama K. Al-Shaykh, Eugene Miloslavsky, Toshio Nomura, Ralph Neff, and Avideh Zakhor. Video compression using matching pursuits. *IEEE Transactions on Circuits and Systems for Video Technology*, 9, February 1999.

[3] A.O. Barut and R. Raczka. *Theory of Group Representations and Applications*. World Scientific Pub Co, 1987.

[4] F. Bergeaud and S. Mallat. Matching pursuit of images. In *Proc. IEEE Int. Conf. Image Proc.*, volume I, pages 53–56, Washington DC, october 1995.

[5] P. Burt and E. Adelson. The laplacian pyramid as a compact image coder. *IEEE Trans. on Comm.*, 31:482–540, 1983.

[6] J. G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Amer*, 2(7):1160–1169.

[7] Geoffrey Davis, Stéphane Mallat, and Marco Avellaneda. Adaptative greedy approximations. *Constructive Approximations*, 1997. Springer-Verlag New YORK INC.

[8] L. Davis, editor. *Handbook of Genetic Algorithms*. Van Nostrand, 1991.

[9] S. Mallat. *A wavelet tour of signal processing*. Academic Press, USA, 1998.

[10] S. Mallat and Zhang. Matching pursuit with time-frequency dictionaries. In *IEEE Transactions on Signal Processing*, number 12, december 1993.

[11] R. Neff and A. Zakhor. Very low bit rate video coding based on matching pursuits. In *IEEE Trans. Circuits and Systems for Video Technology*, volume 7, February 1997.

[12] Bruno A. Olshausen and David J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, (381):607–9, 1996. Springer-Verlag New YORK INC.

[13] Pierre Vandergheynst and Pascal Frossard. Efficient image representation by anisotropic refinement in matching pursuit. In *Proceedings of IEEE*, volume 3, Salt Lake City UT, May 2001. ICASSP.