MATCHING PURSUIT OF IMAGES

François Bergeaud and Stéphane Mallat

Courant Institute of Mathematical Sciences, New York University
251, Mercer Street, New York, NY 10012
Ecole Centrale Paris, Applied Mathematics Laboratory
Grande voie des vignes, F92290 Châtenay-Malabry, France

ABSTRACT

A crucial problem in image analysis is to construct efficient low-level representations of an image, providing precise characterization of features wich compose it, such as edges and texture components.

An image usually contains very different types of features, which have been successfully modelled by the very redundant family of 2D Gabor oriented wavelets, describing the local properties of the image: localization, scale, preferred orientation, amplitude and phase of the discontinuity.

However, this model generates representations of very large size. Instead of decomposing a given image over this whole set of Gabor functions, we use an adaptative algorithm (called Matching Pursuit) to select the Gabor elements wich approximate at best the image, corresponding to the main features of the image .

This produces compact representation in terms of few features that reveal the local image properties. Results proved that the elements are precisely localized on the edges of the images, and give a local decomposition as linear combinations of "textons" in the textured regions.

We introduce a fast algorithm to compute the Matching Pursuit decomposition.

1. INTRODUCTION

The complexity of image structures including different types of textures and edges requires flexible image representations. Although an image is entirely characterized by its decomposition in a basis, any such basis is not rich enough to represent efficiently all potentially interesting low-level structures. Some image components are diffused across many bases elements and are then difficult to analyze from the basis representation. This is like trying to express oneself in a language including a small dictionary. Non available words must be replaced by long awkward sentences. To provide explicit information on important local properties, the image is represented as a sum of waveforms selected from an extremely redundant dictionary of oriented Gabor functions. As opposed to previous approaches, we do not decompose the image over the whole dictionary, but like in a sentence formation, we select the most appropriate Gabor waveforms to represent the image. Instead of increasing the representation by a large factor as in typical multiscale Gabor representations, the adaptive choice of dictionary vectors defines a compact representation that takes advantage of the flexibility offered by the dictionary redundancy.

There is an infinite number of ways to decompose an image over a redundant dictionary of waveforms. The selection of appropriate waveforms to construct the image representation is obtained by constructing efficient image approximations from few dictionary vectors. The optimization of the approximation is not intended for data compression but as a criteria for feature selection. If most of the image is recovered as a sum of few dictionary vectors, these vectors must closely match the local image properties. One can however prove that finding optimal approximations in redundant dictionaries is an NP complete problem. The redundancy opens a combinatorial explosion. This explosion is avoided by the matching pursuit algorithm that uses a non optimal greedy strategy to select each dictionary elements. For a dictionary of Gabor functions, the greedy optimization of the image approximation leads to an efficient image representation where each Gabor waveform reflects the orientation, scale and phase of local image variations. For textures, the selected Gabor elements can be interpreted as textons where as along edges, the multiscale properties of these Gabor elements reflect the edge properties. When the image is translated or rotated, the selected Gabor elements are translated and rotated. A fast implementation of this algorithm and numerical examples are presented.

This work was supported by the AFOSR grant F49620-93-1-0102, ONR grant N00014-91-J-1967 and the Alfred Sloan Foundation

2. THE 2D GABOR WAVELET DICTIONARY

Image decompositions in families of Gabor functions characterize the local scale, orientation and phase of the image variations. Gabor functions are constructed from a window $b(x,y)$, modulated by sinusoidal waves of fixed frequency ω_0 that propagate along different direction θ with two different phases $\phi = 0$ and $\phi = \frac{\pi}{2}$

$$(1) \qquad b_{\theta,\phi}(x,y) = b(x,y)\cos(\omega_0(x\cos\theta + y\sin\theta) + \phi).$$

Each of these modulated windows can be interpreted as wavelets having different orientation selectivities. The window $b(x,y)$ is not chosen to be a Gaussian but is a compactly supported box spline that is adjusted so that the average of $b_{\theta,\phi}(x,y)$ is zero for all orientations and phases.

These oriented wavelets are then scaled by s and translated to define a whole family of Gabor wavelets $\{g_\gamma\}_{\gamma\in\Gamma}$ with:

$$(2) \qquad g_\gamma(x,y) = \frac{1}{s} b_{\theta,\phi}\left(\frac{x-u}{s}, \frac{y-v}{s}\right)$$

where the multi-index parameter $\gamma = (\theta, \phi, s, u, v)$ carries the orientation, phase,scale and position of the corresponding Gabor function.

The Gabor transform of an image $f(x,y)$ is defined by the inner product

$$Gf(\gamma) = < f, g_\gamma > = \iint_{R\times R} f(x,y) g_\gamma(x,y)\, dx\, dy.$$

The Fourier transform of a Gabor function is a waveform whose energy is well concentrated in the Fourier plane. In numerical computations, the scale is restricted to powers of two

0-8186-7310-9/95 $4.00 © 1995 IEEE

$\{2^j\}_{j\in Z}$ and the angles are discretized. The Gabor dictionary used in this paper includes 8 orientations. To define a complete representation, we guaranty that the whole Fourier plane is covered by dilations of the 8 elementary Gabor wavelets.

The decomposition of images in a Gabor dictionary defines a very redundant representation. For an image of 512 by 512 pixels, 8 orientations, 6 octaves and a two phases ($\phi = 0$ and $\phi = \frac{\pi}{2}$) representation would correspond to 96 images of 512 by 512 pixels. These images could be subsampled but we then lose the translation invariance of the representation. Instead of decomposing the image over the whole dictionary, we select specific Gabor waveforms that provide an efficient image approximation.

3. MATCHING PURSUIT

We consider the general problem of decomposing a signal f over a dictionary of unit vectors $\{g_\gamma\}_{\gamma\in\Gamma}$ whose linear combinations are dense in the signal space $\mathcal{H}$. The smallest possible dictionary is a basis of $\mathcal{H}$; general dictionaries are redundant families of vectors. When the dictionary is redundant, unlike the case of a basis, we have some degree of freedom in choosing a signal's particular representation. This freedom allows us to choose few dictionary vectors, whose linear combinations approximate efficiently the signal. The chosen vectors highlights the predominant signal features. For any fixed approximation error ϵ, when the dictionary is redundant, we can show that finding the minimum number of dictionary elements that approximates the image with an error smaller than ϵ is an NP hard problem.

Because of the difficulty of finding optimal solutions, we use a greedy matching pursuit algorithm that has previously been tested on a one-dimensional signal.The matching pursuit ([3]) uses a greedy strategy that computes a good suboptimal approximation. It successively approximates a signal f with orthogonal projections onto dictionary elements.

Let $R^0 f = f$. Suppose that we have already computed the residue $R^k f$.

We choose $g_{\gamma_k} \in \mathcal{D}$ such that:

$$|< R^k f, g_{\gamma_k} >| = \sup_{\gamma\in\Gamma} |< R^k f, g_\gamma >|. \tag{3}$$

and project $R^k f$ on g_{γ_k}

$$R^{k+1} f = R^k f - < R^n k f, g_{\gamma_k} > g_{\gamma_k}. \tag{4}$$

which defines the residue at the order k+1. The orthogonality of $R^{k+1} f$ and g_{γ_k} implies

$$\|R^{k+1} f\|^2 = \|R^k f\|^2 - |< R^k f, g_{\gamma_k} >|^2. \tag{5}$$

By summing (4) for k between 0 and $n-1$, we obtain

$$f = \sum_{k=0}^{n-1} < R^k f, g_{\gamma_k} > g_{\gamma_k} + R^n f. \tag{6}$$

The residue $R^n f$ is the approximation error of f after choosing n vectors in the dictionary.

In infinite dimensional spaces, the convergence of the error to zero is shown ([3]) to be a consequence of a theorem proved by Jones ([2]):

$$f = \sum_{n=0}^{+\infty} < R^n f, g_{\gamma_n} > g_{\gamma_n}, \tag{7}$$

and we obtain an energy conservation

$$\|f\|^2 = \sum_{n=0}^{+\infty} |< R^n f, g_{\gamma_n} >|^2. \tag{8}$$

In finite dimensional signal spaces, the convergence is proved to be exponential ([3]).

4. RESULTS

The matching pursuit algorithm applied to a Gabor dictionary selects iteratively the Gabor waveforms, also called atoms, whose scales, phases, orientations and positions best match the local image variations.

In order to display the edge information (localization, orientation, scale, amplitude), we adopt the following convention: each selected Gabor vector g_γ for $\gamma_n = (\theta, \phi, 2^j, u, v)$ is symbolized by an elongated Gaussian function of width proportional to the scale 2^j, centered at (u, v), of orientation θ. The mean gray level of each symbol is proportional to $|< R^n f, g_{\gamma_n} >|$.

Figure (4) displays the reconstruction of the Lena image with the 2500 first Gabor waveforms selected by the algorithm. The image reconstruction is simply obtained with the truncated sum of the infinite serie decomposition (7). With 2500 atoms, we already recover a very good image quality and with 5000 atoms the reconstructed image has no visual degradation. The latter image is constructed without using any compression method, and achieves a 1.5 bit/pixel ratio with a perfect visual quality. This shows clearly that, while specifying precisely the important information such as the local scale and orientation, this representation is very compact and can be used as the input to a high level processing. Moreover, the degradation of the image quality is very small when using 2500 atoms instead of 5000 atoms, which suggests that the most relevant features of the image lie in a very small set of atoms.

The figures 6, 7, and 8 illustrate the texture discrimination properties of the representation: a straw texture image is inserted into a paper texture image. The paper texture has no orientation specificity. On the other hand, the straw texture has horizontal and vertical structures. At fine scales, most structures are vertical because the horizontal variations are relatively smooth. At the intermediate scale 2^2 we clearly see the horizontal and vertical image structures. At the larger scale 2^3, the vertical structures dominate again because the vertical straw are of larger sizes.

The symbolic representations at scale 2^1, 2^2, 2^3, show the distinct behaviour of the two textures relative to the scale: although the paper texture shows a scale-invariant uniform distribution, the straw texture representation points out the vertical features of the straw at finer and larger scale, and the horizontal and vertical structures of the straw at medium scale. Moreover, at scale 2^1, we clearly distinguish the texture edge as the boundary between the two textures.

In the example of the straw texture inserted into a paper texture, the discrimination is achieved using the information concerning the textons, and mainly on the energy density relative to the orientation and the scale. Indeed, the energy distribution of the paper is nearly constant at a given scale for different orientations, and is concentrated in the vertical "cells" at low and high scales and in the vertical and horizontal "cells" at intermediate scales.

Figure 1: Symbolic representation of the Gabor vectors at scale 2^1: each vector is symbolized by an elongated Gaussian of width proportional to the scale 2^j, oriented along the modulation direction, of gray level proportional to the amplitude $|< R^n f, g_{\gamma_n} >|$. Light atoms correspond to high amplitude.

Figure 2: Symbolic representation of the Gabor vectors at scale 2^2.

Figure 3: Original Lena image (256×256)

Figure 4: Reconstruction with 2500 atoms. The reconstruction is obtained as a truncated sum of the decomposition.

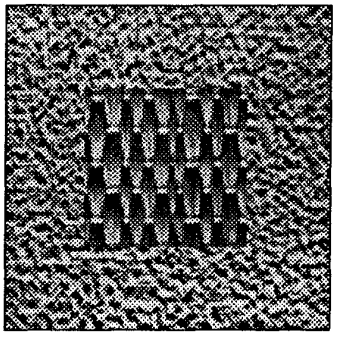

Figure 5: Original paper and straw texture mosaic (256×256)

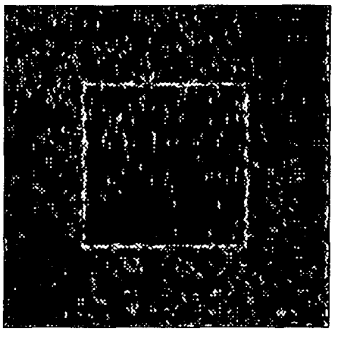

Figure 6: Symbolic representation of the Gabor vectors for the texture mosaic: scale 2^1.

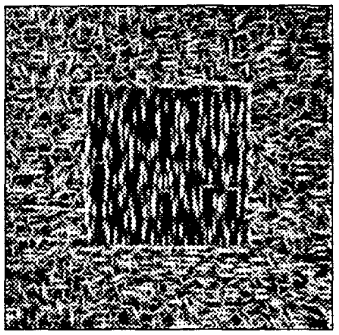

Figure 7: Symbolic representation of the Gabor vectors for the texture mosaic: scale 2^2.

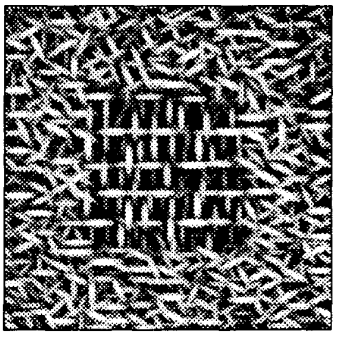

Figure 8: Symbolic representation of the Gabor vectors for the texture mosaic: scale 2^3.

5. FAST NUMERICAL COMPUTATIONS

At a first glance, a matching pursuit seems to require a hopeless amount of computations. These computations can however be considerably reduced with an efficient algorithm that prunes the dictionary by removing the elements which are *a priori* not significant. Indeed, the brute force algorithm is very slow, because it deals with dictionary vectors which do not correspond to edges or texture elements. Moreover those features necessarily belongs to the local maxima of the coefficients along the directions of the filters. We thus compute a sub-dictionary adapted to the image by selecting in the initial dictionary the vectors correponding to a local maxima of the representation.

For $f \in \mathcal{H}$, we call a local maxima in the parameter space Γ an index γ_0 such that for all γ in a neighborhood of γ_0 in Γ

$$| < f, g_\gamma > | \leq | < f, g_{\gamma_0} > |. \tag{9}$$

For example, in a wavelet dictionary of two-dimensional oriented atoms, the local maxima are computed for fixed scale and for fixed direction. For each scale s and direction θ_0, the local maxima are defined as indexes $\gamma_0 = (\theta_0, \phi_0, s, u_0, v_0)$ such that (9) is valid for any $\gamma = (\theta_0, \phi, s, u, v)$ with (u, v) in a monodimentional oriented (direction θ_0) neighborhood of (u_0, v_0).

At the step 1 of the algorithm we prune the dictionary with a local maxima selection. All inner products $\{< f, g_\gamma >\}_{\gamma \in \Gamma}$ are computed. We choose a threshold ϵ and select only the local maxima that are large enough

$$| < f, g_\gamma > | \geq \epsilon \|f\|.$$

The matching pursuit is then computed by induction as follow.

Suppose that the first n vectors $\{g_{\gamma_k}\}_{0 \leq k < n}$ have been selected. We denote by Γ_n the indexes γ such that $| < f, g_\gamma > |$ is a local maxima and $| < R^n f, g_{\gamma_0} > | \geq \epsilon \|f\|$. We find g_{γ_n} which correlates $R^n f$ at best in this reduced dictionary

$$| < R^n f, g_{\gamma_n} > | = \sup_{\gamma \in \Gamma_n} | < R^n f, g_\gamma > |.$$

We compute the inner product of the new residue $R^{n+1} f$ with all $\{g_\gamma\}_{\gamma \in \Gamma_n}$ with an updating formula derived from equation (4)

$$< R^{n+1} f, g_\gamma > = < R^n f, g_\gamma > - < R^n f, g_{\gamma_n} > < g_{\gamma_n}, g_\gamma > . \tag{10}$$

Since we previously stored $< R^n f, g_\gamma >$ and $< R^n f, g_{\gamma_n} >$, this update is obtained in O(1) operations if the value $< g_{\gamma_n}, g_\gamma >$ can be retrieved in $O(1)$ operations, which is the case for the dictionary of two-dimensional wavelets. The vectors in these dictionaries have a sparse interaction which means that for most $\gamma \in \Gamma_n$, we have $< g_{\gamma_n}, g_\gamma >= 0$. There are thus few indexes γ for which the value of $< R^n f, g_\gamma >$ must be updated. The dictionary is further pruned by suppressing from Γ_n all indexes γ such that $| < R^{n+1} f, g_\gamma > | < \epsilon \|f\|$. The iteration is then continued on this new index set Γ_{n+1}.

If we iterate this procedure, the index Γ_n is progressively reduced until it gets empty for $n = m$. We then come back to the step 1 and replace f by $R^m f$. The local maxima of $< R^m f, g_\gamma >$ are computed and are thresholded with the new value $\epsilon \|R^m f\|$. The pursuit is then continued on these maxima with the iteration previously described, until the index set is again empty for $n = p$. We come back again to step 1 by replacing f by $R^p f$ and continue the iterations.

6. VISION APPLICATIONS

We introduced here a method to construct a decomposition of images into its main features. We showed that this transform provides a precise and complete characterization of the edges and texture components in terms of localization, orientation, scale and amplitude. By reconstructing high-visual quality images with very few atoms, we also showed that this representation is compact.

Another advantage of Matching Pursuit is the flexibility of the dictionary choice allowing to explicitely introduce *a priori* knowledge on the features of object classes into the dictionary to solve specific vision problems.

7. REFERENCES

[1] J. Daugman, "Complete discrete 2D Gabor transform by neural networks for image analysis and compression", *IEEE Trans. on Acoustic, Speech, and Signal Processing*, ASSP-36, pp 1169-1179, 1988.

[2] L. K. Jones, "On a conjecture of Huber concerning the convergence of projection pursuit regression", *The Annals of Statistics*, vol. 15, No. 2, pp. 880-882, 1987.

[3] S. Mallat and Z. Zhang, "Matching Pursuit with time-frequency dictionaries", *IEEE Trans. on Signal Processing*, Dec. 1993.

[4] J. Malik and P. Perona, "Preattentive texture discrimination with early vision mechanisms", *J. Opt. Soc. Am.*, vol. 7, No. 5, pp. 923-932, May 1990.

[5] S. Marcelja, "Mathematical description of the response of simple cortical cells", *J. Opt. Soc. Am.*, vol. 70, No. 11, pp 1297-1300, November 1980.