Processing images and sounds with matching pursuits

François Bergeaud and Stéphane Mallat

Courant Institute of Mathematical Sciences
Computer Science Department
New York University
251, Mercer Street
New York, NY 10012, USA

ABSTRACT

Complex signals need to be adaptively expanded over families of waveforms selected to match their different structures. A matching pursuit is a greedy algorithm that expands signals over vectors that are selected among redundant dictionaries of waveforms. The properties of this algorithms are reviewed. Applications to sound and image processing with dictionaries of time-frequency atoms are described.

Keywords: low-level representations, wavelets, matching pursuit, Gabor decomposition, adaptative algorithm, texture discrimination, early vision, sound processing

1 INTRODUCTION

The complexity of structures encountered in a sound recording or an image requires to develop adaptive low-level representations. Although such a signal is entirely characterized by its decomposition in a basis, a basis is a minimal set of vectors that is not rich enough to represent efficiently all components. Some signal structures are diffused across many basis elements and are thus difficult to analyze from this expansion. For example, image variations corresponding to edges and textures are not efficiently represented by the same types of waveforms. The same issue appears in sound recordings that includes transients that are well represented by short waveforms, and harmonics that are more efficiently decomposed over long waveforms with short frequency support. Instead of decomposing all signals over the same family of waveforms, adaptive transforms choose the decomposition vectors depending upon the signal properties. These vectors are selected among a family of waveforms that is much larger than a basis, which is called a dictionary.

There is an infinite number of ways to decompose a signal over a redundant dictionary of waveforms. Our goal is to find an expansion that can efficiently approximate the signal with few dictionary vectors. This optimization is not only intended for data compression but also as a criteria for feature selection. If most of the signal is recovered as a sum of few dictionary vectors, these vectors can be interpreted as essential signal components. However, the dictionary redundancy opens a combinatorial explosion and one can prove that finding optimal approximations is NP hard. This computational explosion is avoided with a suboptimal greedy algorithm called matching pursuit. The main properties of this approximation algorithm is reviewed in this article. Applications to sound processing with a dictionary of local time-frequency atoms are studied in section 3. For image processing, the dictionary is

0-8194-1928-1/95/$6.00

composed of wavelets with several orientation tunings, that respond to edges and texture variations.

2 MATCHING PURSUIT

A dictionary is a family of vectors $\mathcal{D} = (g_\gamma)_{\gamma \in \Gamma}$ included in a Hilbert space H, with a unit norm $||g_\gamma|| = 1$. We suppose that the dictionary is complete in H which means that linear combinations of dictionary vectors are dense in H. The smallest possible dictionaries are bases of H. To implement adaptive expansions, we rather consider much larger and very redundant dictionaries. The redundancy allows one to choose from the dictionary the vectors that are best adapt to expand any particular vector in H.

Let $f \in H$. For any error $\epsilon > 0$, one may define an optimal approximation in $\mathcal{D}$ as an expansion over a minimum number M of dictionary vectors $\{g_{\gamma_n}\}_{0 \leq n < M}$ in $\mathcal{D}$ such that

$$\|f - \sum_{0 \leq n < M} \beta_n g_{\gamma_n}\| \leq \epsilon.$$

One can prove that for general dictionaries, finding this optimal expansion is an NP-hard problem.[2]

To avoid this combinatorial explosion, a matching pursuit uses a suboptimal greedy strategy. The vector f is successively approximated with orthogonal projections onto vectors that are chosen in $\mathcal{D}$ to minimize the residual errors. Let $g_{\gamma_0} \in \mathcal{D}$. The vector f can be decomposed into

$$f = < f, g_{\gamma_0} > g_{\gamma_0} + Rf, \tag{1}$$

where Rf is the residual vector after approximating f in the direction of g_{γ_0}. Clearly g_{γ_0} is orthogonal to Rf, hence

$$\|f\|^2 = |< f, g_{\gamma_0} >|^2 + \|Rf\|^2. \tag{2}$$

To minimize $\|Rf\|$, we must choose $g_{\gamma_0} \in \mathcal{D}$ such that $|< f, g_{\gamma_0} >|$ is maximal. In some cases, it is only possible to find a vector g_{γ_0} that is close to the maximum in the sense that

$$|< f, g_{\gamma_0} >| \geq \alpha \sup_{\gamma \in \Gamma} |< f, g_\gamma >|, \tag{3}$$

where $\alpha \in (0, 1]$ is an optimality factor.

We sub-decompose the residue Rf by projecting it onto the vector of $\mathcal{D}$ that best matches Rf, as it was done for f. This projection of Rf generates a second residue, $R^2 f$, which we again decompose to obtain a third residue, and so on.

Let us explain by induction how the matching pursuit is carried further. Let $R^0 f = f$. We suppose that we have computed the n^{th} order residue $R^n f$, for $n \geq 0$. We choose an element $g_{\gamma_n} \in \mathcal{D}$ which closely matches the residue $R^n f$ in the sense that

$$|< R^n f, g_{\gamma_n} >| \geq \alpha \sup_{\gamma \in \Gamma} |< R^n f, g_\gamma >|. \tag{4}$$

The residue $R^n f$ is sub-decomposed into

$$R^n f = < R^n f, g_{\gamma_n} > g_{\gamma_n} + R^{n+1} f, \tag{5}$$

which defines the residue at the order n+1. Since $R^{n+1} f$ is orthogonal to g_{γ_n}, we have

$$\|R^n f\|^2 = |< R^n f, g_{\gamma_n} >|^2 + \|R^{n+1} f\|^2. \tag{6}$$

By iterating this decomposition up to the order m, we can decompose f into the telescoping sum

$$f = \sum_{n=0}^{m-1} (R^n f - R^{n+1} f) + R^m f. \tag{7}$$

Equation (5) yields

$$f = \sum_{n=0}^{m-1} < R^n f, g_{\gamma_n} > g_{\gamma_n} + R^m f. \tag{8}$$

Similarly, we write $\|f\|^2$ as a telescoping sum

$$\|f\|^2 = \sum_{n=0}^{m-1} (\|R^n f\|^2 - \|R^{n+1} f\|^2) + \|R^m f\|^2 \tag{9}$$

which we combine with (6) to obtain an energy conservation equation

$$\|f\|^2 = \sum_{n=0}^{m-1} |< R^n f, g_{\gamma_n} >|^2 + \|R^m f\|^2. \tag{10}$$

A matching pursuit decompose any f into a sum of dictionary elements which are chosen to best match its residues. Although this decomposition is non-linear, we maintain an energy conservation as though it was a linear, orthogonal decomposition. An important issue is to understand the behavior of the residue $R^m f$ when m increases. By transposing a result proved by Jones[5] for projection pursuit algorithms,[3] it has been proved[6] that the matching pursuit converges, even in infinite dimensional spaces.

THEOREM 2.1. *Let $f \in H$. The residue $R^m f$ defined by the induction equation (5) satisfies*

$$\lim_{m \to +\infty} \|R^m f\| = 0. \tag{11}$$

Hence

$$f = \sum_{n=0}^{+\infty} < R^n f, g_{\gamma_n} > g_{\gamma_n}, \tag{12}$$

and

$$\|f\|^2 = \sum_{n=0}^{+\infty} |< R^n f, g_{\gamma_n} >|^2. \tag{13}$$

When H is of finite dimension, $||R^m f||$ decays exponentially to zero.

This theorem proves that any vector f is characterized by the double sequence $(< R^n f, g_{\gamma_n} >, \gamma_n)_{n \in N}$, which specifies the expansion coefficients and the index of each chosen vector within the dictionary. After m iterations, (8) shows that the approximation error is

$$R^m f = f - \sum_{n=0}^{m-1} < R^n f, g_{\gamma_n} > g_{\gamma_n}. \tag{14}$$

The best approximation of f as a a linear expansion of $\{g_{\gamma_n}\}_{0 \leq n < m}$ is the orthogonal projection of f on the space generated by this family of vectors. In general the vectors $\{g_{\gamma_n}\}_{0 \leq n < m}$ are not orthogonal so the matching pursuit expansion is not equal to the orthogonal projection of f. An improved approximation can be recovered by orthogonalizing the family $\{g_{\gamma_n}\}_{0 \leq n < m}$ with a Gram-Schmidt procedure and computing the orthogonal projection of f,[2,7] Such an orthogonal pursuit gives the better approximations at the cost of an increase computational complexity. In general, the improved approximation obtained with an orthogonal pursuit is not worth the computational expenses, in large dimensional spaces.

The asymptotic convergence of a matching pursuit has further been studied by analyzing the behavior of the normalized residue

$$\tilde{R}^n f = \frac{R^n f}{\|R^n f\|}.$$

The non-linear map defined by $\tilde{R}^{n+1} f = \mathcal{M}(\tilde{R}^n f)$ exhibits chaotic properties. Experimental data suggest that the normalized residues of a normalized pursuit converge to a chaotic attractor. In low-dimensional spaces, it has been proved that $\mathcal{M}$ is topologically equivalent to a left-shift map operator,[2] whose chaotic properties are entirely known. In high dimensional spaces, the analysis was performed for a particular dictionary of Diracs and complex exponentials. Numerically, one can observe that the first few iterations of the pursuit extracts the components of f which are strongly correlated with dictionary vectors, which we call coherent part. The remaining residue does not correlate strongly to any dictionary vectors and its properties depend upon the attractor of the chaotic map. For dictionaries of time-frequency atoms described in the next section, these residues converge to realizations of white noises.[2] It is thus often not worth it to further reduce the energy of these residues by continuing the pursuit.

3 SOUND PURSUIT

To analyze the time and frequency localization properties of one-dimensional signals such as speech or music recordings, we use a large dictionary of time-frequency atoms. Our signal space is $L^2(R)$ and we construct such a dictionary by scaling, translating and modulating a single window function $g(t) \in L^2(R)$. We suppose that $g(t)$ is an even and real function of unit norm. For any scale $s > 0$, frequency modulation ξ and translation u, we denote $\gamma = (s, u, \xi)$ and define

$$g_\gamma(t) = \frac{1}{\sqrt{s}} g(\frac{t-u}{s}) e^{i\xi t}. \tag{15}$$

The index γ is an element of the set $\Gamma = R^+ \times R^2$. The factor $\frac{1}{\sqrt{s}}$ normalizes to 1 the norm of $g_\gamma(t)$. The function $g_\gamma(t)$ is centered at the abscissa u and its energy is concentrated in a neighborhood of u, whose size is proportional to s. Let $\hat{g}(\omega)$ be the Fourier transform of $g(t)$. Equation (15) yields

$$\hat{g}_\gamma(\omega) = \sqrt{s}\hat{g}(s(\omega - \xi)) e^{-i(\omega-\xi)u}. \tag{16}$$

Since $|\hat{g}(\omega)|$ is even, $|\hat{g}_\gamma(\omega)|$ is centered at the frequency $\omega = \xi$. Its energy is concentrated in a neighborhood of ξ, whose size is proportional to $\frac{1}{s}$. The dictionary of time-frequency atoms $\mathcal{D} = \{g_\gamma(t)\}_{\gamma \in \Gamma}$ is a very redundant set of functions in $L^2(R)$ that includes window Fourier frames and wavelet frames.

A matching pursuit chooses the time-frequency atoms of $\mathcal{D}$ which are "best" adapted to expand f. Since a time-frequency atom dictionary is complete, Theorem 2.1 proves that

$$f = \sum_{n=0}^{+\infty} < R^n f, g_{\gamma_n} > g_{\gamma_n}, \tag{17}$$

where $\gamma_n = (s_n, u_n, \xi_n)$ and

$$g_{\gamma_n}(t) = \frac{1}{\sqrt{s_n}} g(\frac{t-u_n}{s_n}) e^{i\xi_n t}. \tag{18}$$

Each atom $g_{\gamma_n}(t)$ can be represented by its Wigner distribution which specifies its localization in time and frequency

$$Wg_{\gamma_n}(t,\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} g_{\gamma_n}(t + \frac{\tau}{2}) g^*_{\gamma_n}(t - \frac{\tau}{2}) e^{-i\omega\tau} d\tau. \tag{19}$$

If $g(t)$ is the Gaussian window

$$g(t) = 2^{1/4} e^{-\pi t^2}, \tag{20}$$

the time-frequency atoms $g_\gamma(t)$ are called Gabor functions. One can verify that

$$Wg_\gamma(t,\omega) = |g_\gamma(t)|^2 |\hat{g}_\gamma(\omega)|^2 = 2e^{-2\pi \frac{(t-u)^2}{s^2}} e^{-\frac{s^2(\omega-\xi)^2}{2\pi}}. \tag{21}$$

It is a two-dimensional Gaussian blob in the time-frequency plane which indicate the energy distribution of f in time and frequency. centered at (u,ξ) whose width in time and frequency is proportional respectively to s and $\frac{1}{s}$.

The time-frequency energy distribution of $f(t)$ is defined by

$$Ef(t,\omega) = \sum_{n=0}^{+\infty} |< R^n f, g_{\gamma_n} >|^2 W g_{\gamma_n}(t,\omega). \tag{22}$$

It is thus a sum of two-dimensional Gaussian blob in the time-frequency plane. From the energy conservation equation (13), one can verify a time-frequency energy conservation

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} Ef(t,\omega) dt\, d\omega = \sum_{n=0}^{+\infty} |< R^n f, g_{\gamma_n} >|^2 = \|f\|^2. \tag{23}$$

Figure 5 is the graph of a speech recording corresponding to the words "Matching Pursuit", sampled at 8 kHz. From the time-frequency energy displayed in Figure 6, we can distinguish the syllables "ma", "tch", "ing", "pur", "s", "uit". They are represented as different kind of atom aggregations in the time-frequency plane. The syllable "ma" consists in low-frequency harmonics and diracs. The "tch" and "s" have energy spread over medium and high frequencies. The "ing" is composed of low-frequency gabor atoms and diracs. Moreover, the letters "t" and "p" are easily revealed as strong isolated diracs.

For $n = 500$ atoms, $\frac{\|R^n f\|}{\|f\|} = 0.118$, although the signal has 8192 samples, the sound recovered from these atoms is of excellent quality. For a dictionary of Gabor atoms, the matching pursuit residue converges to realizations of white noises. The residue $R^n f$ displayed in Figure 3 is already close to such a realization. The first 500 atoms thus correspond to the coherent part of the sound with respect to our time-frequency dictionary. The extraction of coherent parts has applications to noise removal, which are studied in,[2].[6]

4 IMAGE PURSUIT

For image processing, we must select a dictionary that can characterize the local scale and orientation of the image variations. For this purpose, we use a dictionary composed of several two-dimensional wavelets that have specific orientation selectivity. These wavelets are all derived from a two-dimensional window $g(x,y)$ that is modulated at a fixed frequency ω_0 along several directions specified by an angle θ in the (x,y) plane

$$g_\theta(t) = g(x,y) \exp[i(x\cos\theta + y\sin\theta)].$$

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

$$g_\gamma(x,y) = \frac{1}{s} g_\theta(\frac{x-u}{s}, \frac{y-v}{s}). \tag{24}$$

The multi-parameters index $\gamma = (\theta, s, u, v)$ carries the orientation, scale and position of the corresponding wavelet.

In numerical computations, the scale is restricted $\{2^j\}_{j\in Z}$ and the angles are discretized. This wavelet dictionary used in this paper includes 8 orientations (cf. figure (1)). To define a complete representation, we guaranty that the whole Fourier plane is covered by dilations of these 8 wavelets.

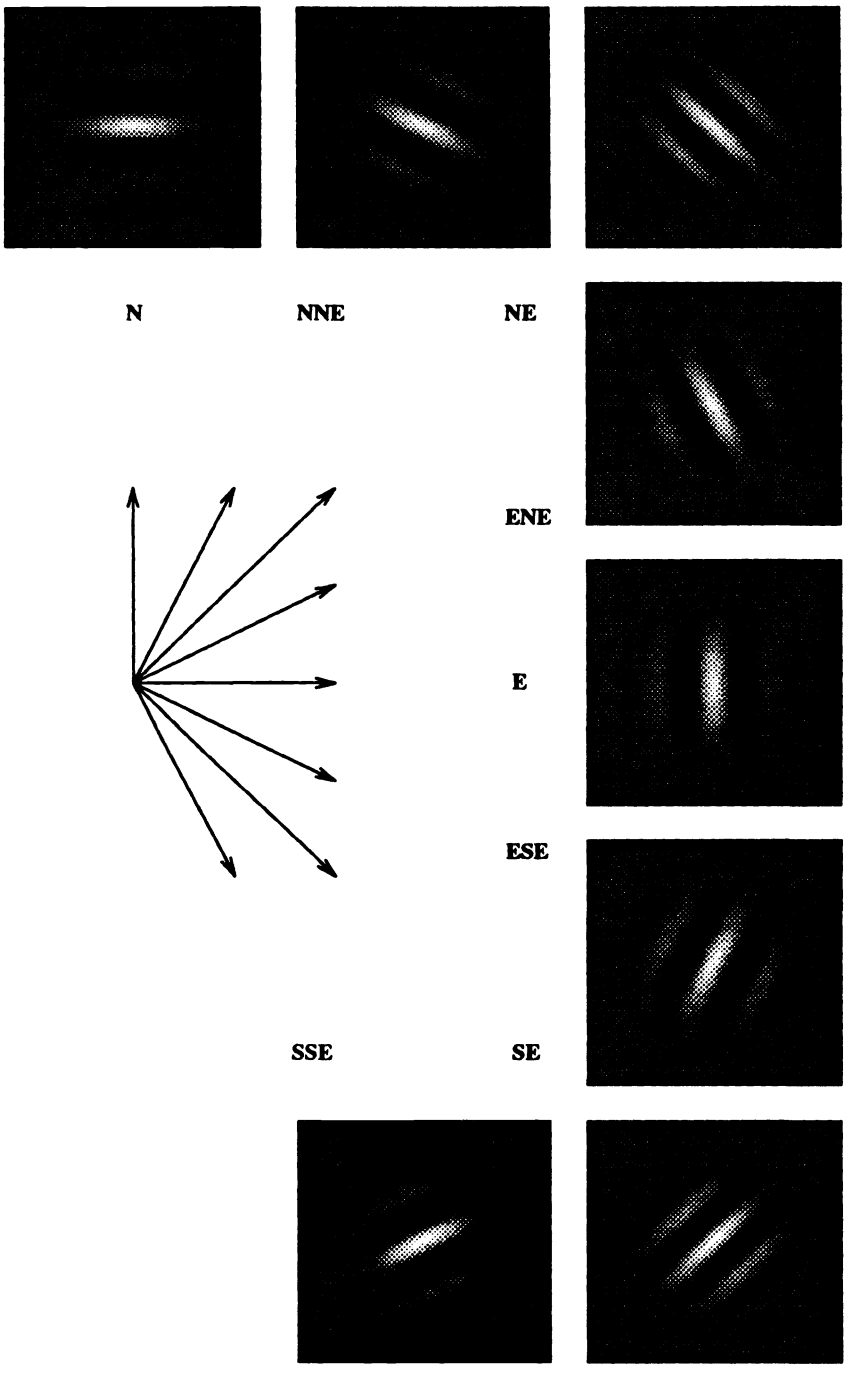

Figure 1: Discretization of the angle θ

Besides engineering and mathematical motivations, families such wavelets have been used as models for the receptive fields of simple cells. In the cat's visual cortex, Hubel and Wiesel[4] discovered a class of cells, called simple cells, whose response depends upon the frequency and orientation of the visual stimuli. Numerous physiological experiments[8] have shown that their response can be modeled with linear filters, whose impulse response have been measured at different locations of the visual cortex. Daugmann[1] showed that these impulse response can be approximated by Gaussian windows modulated by a sinusoidal wave. Depending upon the cortical cell, this modulated Gaussian is dilated and has a specific spatial orientation tuning. These findings suggest the existence of some sort of wavelet transform in the visual cortex, combined with subsequent non-linearities. The frequency resolution of these "physiological" wavelets seems to be of the order of 1-1.5 octaves.

The matching pursuit algorithm applied to this wavelet dictionary selects iteratively the wavelets, whose scales, orientations and positions best match the local image variations. The distribution of wavelets at the scales 2^j for $1 \leq j \leq 4$ is shown in fig. (2). Each selected Gabor vector g_γ for $\gamma = (\theta, 2^j, u, v)$ 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 $|< R^n f, g_{\gamma_n} >|$. Bright atoms correspond to high amplitude inner products. As clearly shown by these images, the fine scale atoms are distributed along the edges and the texture regions and their orientation indicate the local orientation of the image transitions.

Figure 3 displays the reconstruction of the Lena image with the first 2500 wavelets selected by the matching pursuit. With 2500 atoms, the reconstructed image is already of good visual quality, although it is recovered with 25 times less atoms than the number of pixels in the image. These approximations in a redundant wavelet dictionary can be compared with expansion in a minimum dictionary of orthogonal wavelets. With a dictionary of separable wavelets computed with Haar filters, we need 5000 orthogonal wavelets to reconstruct an image with has the same error than the image reconstructed with 1500 wavelets selected from our redundant dictionary. This shows that the freedom of choice offered by the dictionary redundancy allows one to choose wavelets that better match the local image properties.

Figure 4 illustrates the texture discrimination properties of the representation. A texture of straw is inserted into a texture of paper. The paper texture has no orientation specificity. On the other hand, the straw texture has horizontal and vertical structures. The wavelets selected by the matching pursuit are displayed in Fig. 4. The boundary of the two textures is covered by wavelets at the finest scale 2^1. The distribution of wavelets corresponding to the paper texture has no preferential orientations and is self-similar across scales. The energy distribution across scales of this texture is derived from the energy density of wavelets at the different scales. For the straw texture, at fine scales most wavelets are vertical. Indeed the sharpest variations of this texture are along vertical lines. At the intermediate scale 2^2, the selected wavelets correspond to the horizontal and vertical components of the texture. At the larger scale 2^3, there are mainly vertical wavelets because the vertical straw have a larger size. This example illustrates the ability to discriminate textures from the parameters of the wavelets selected by a matching pursuit. An automatic texture segmentation algorithm is currently being developed.

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 with local maxima. For $f \in H$, 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

$$|< f, g_\gamma >| \leq |< f, g_{\gamma_0} >|. \tag{25}$$

For example, in a Gabor dictionary of one-dimensional time frequency atoms, the local maxima are computed for fixed scale. For each scale s, the local maxima are defined as indexes $\gamma_0 = (s, u_0, \xi_0)$ such that (25) is valid for any $\gamma = (s, u, \xi)$ with (u, ξ) in a neighborhood of (u_0, ξ_0).

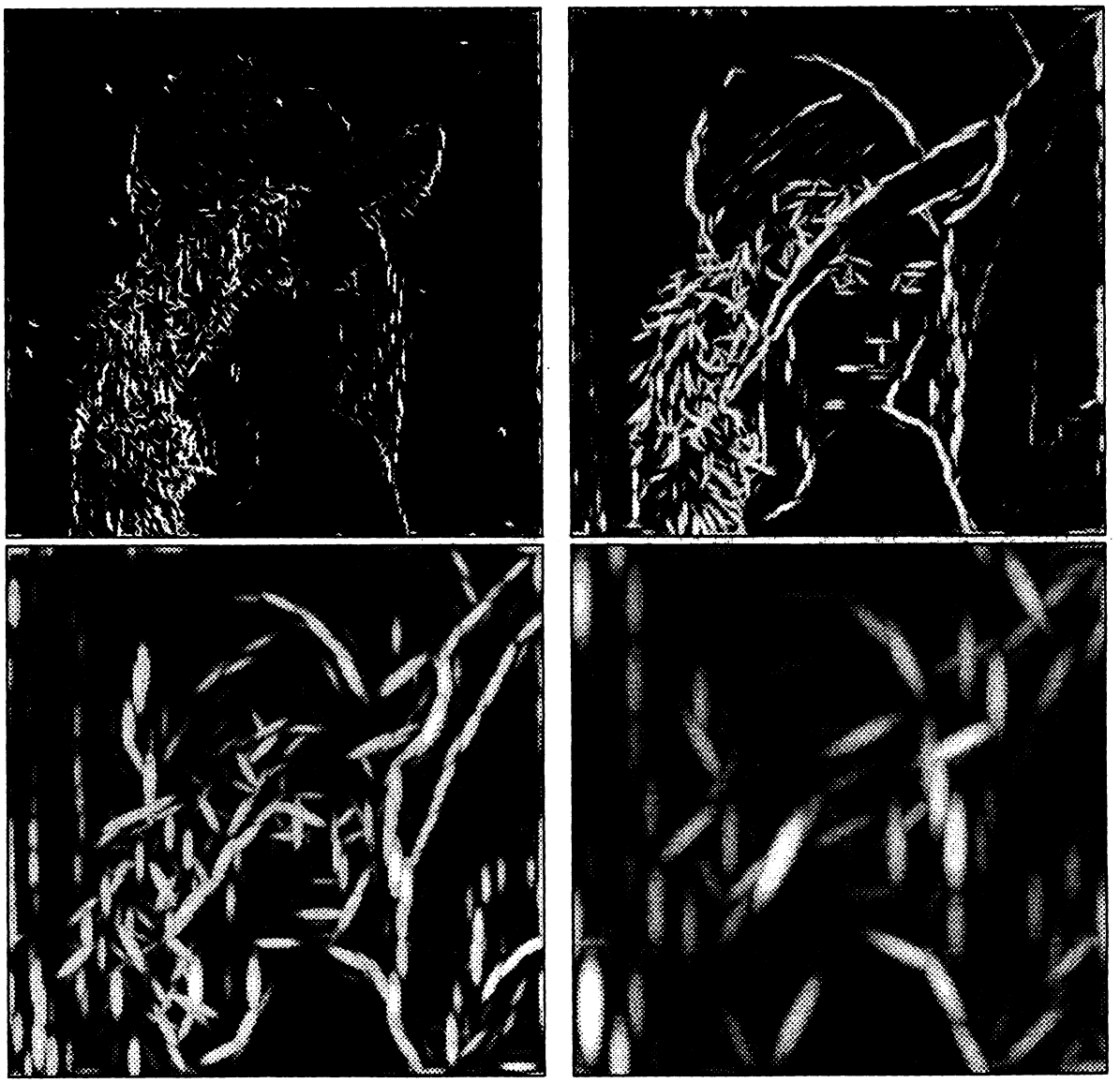

Figure 2: Symbolic representation of the Gabor vectors: each atom is symbolized by an elongated Gaussian of width proportional to the scale 2^j, oriented along the modulation direction, of gray level proportional to its amplitude. Light atoms correspond to high amplitude. The images from left to right and from the top to the bottom corresponds to the scale 2^1 to 2^4.

Figure 3: Original Lena image (left figure) and reconstruction with 2500 (right figure). The reconstruction is obtained as a truncated sum of the decomposition.

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 (5)

$$< R^{n+1} f, g_\gamma > = < R^n f, g_\gamma > - < R^n f, g_{\gamma_n} > < g_{\gamma_n}, g_\gamma > . \tag{26}$$

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. This is the case for the Gabor dictionary of one-dimensional time-frequency atoms and 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. A software implementing matching pursuit for time-frequency dictionaries is available through anonymous ftp at the address cs.nyu.edu . Instructions are in the file README of the directory /pub/wave/software.

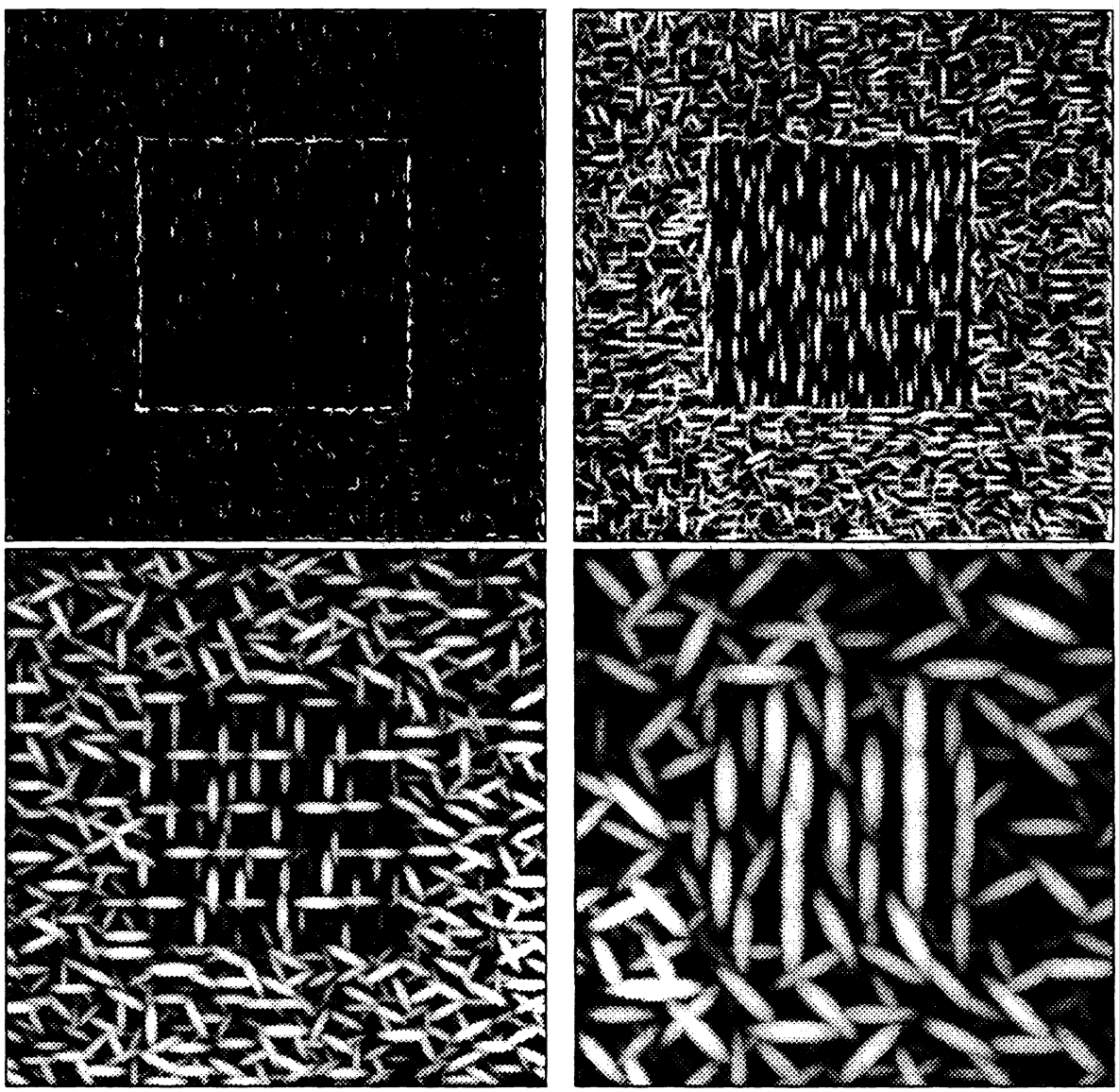

Figure 4: Symbolic representation of the Gabor vectors for the texture mosaic: scale 2^1 (upper-left image), 2^2 (upper-right image), scale 2^3 (down-left image) and 2^4 (down-right image).

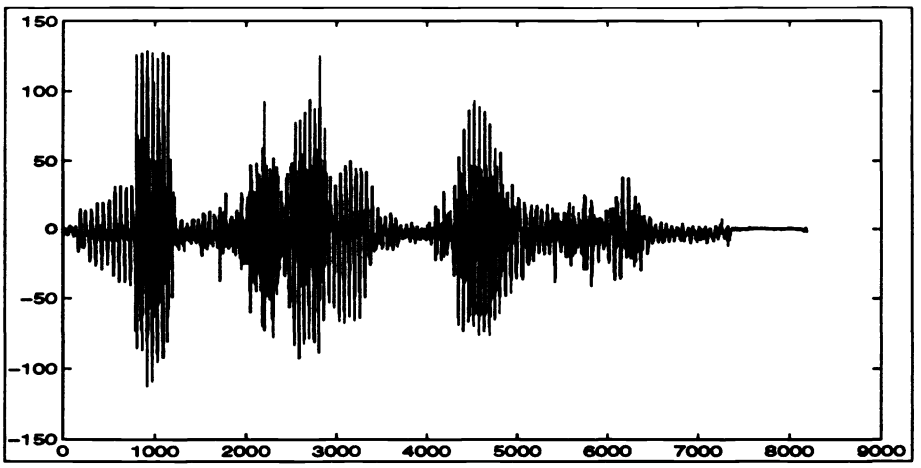

Figure 5: Speech recording of the words "matching pursuit", sampled at 8 kHz.

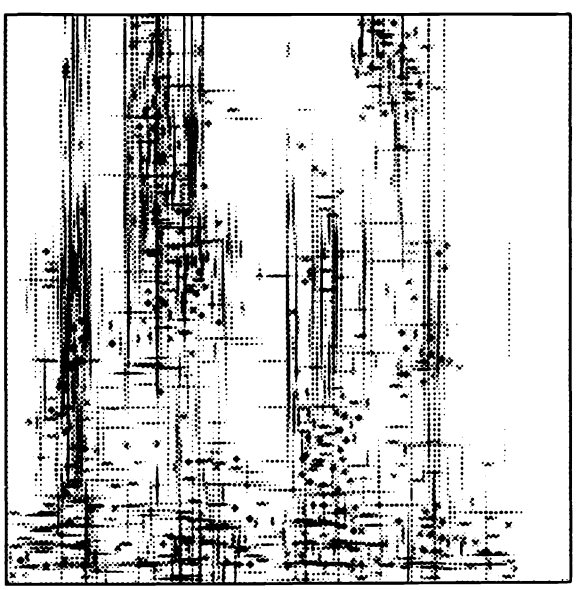

Figure 6: Time-frequency energy distribution of the speech recording. We distinguish the syllables "ma", "tch", "ing", "pur", "s", "uit". They are represented as different kind of atom aggregations in the time-frequency plane. The syllable "ma" consists in low-frequency harmonics and diracs. The "tch" and "s" have energy spread over medium and high frequencies. The "ing" is composed of low-frequency gabor atoms and diracs. Moreover, the letters "t" and "p" are easily revealed as strong isolated diracs.

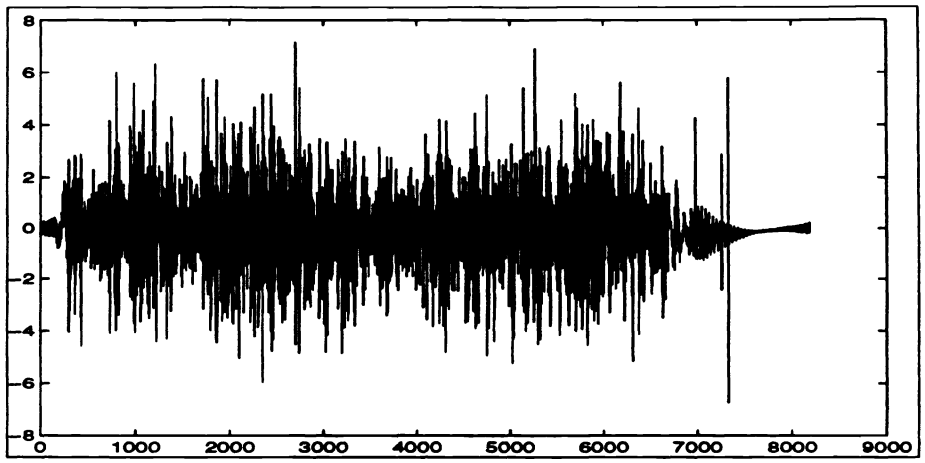

Figure 7: Residue $R^n f$ after $n = 750$ iterations. This residue is very similar to the realization of a white noise.

6 ACKNOWLEDGMENTS

This work was supported by the AFOSR grant F49620-93-1-0102 and the ONR grant N00014-91-J-1967.

REFERENCES

[1] J. G. Daugman, "Two-dimensional spectral analysis of cortical receptive field profile", *Vision Research*, vol.20, pp. 847-856, 1980.

[2] G. Davis, S. Mallat, M. Avellaneda, "Adaptive Nonlinear Approximations", To appear in *Journal of Constructive Approximation.*

[3] J. H. Friedman and W. Stuetzle, "Projection pursuit regression," Journal of the American Statistical Association, Vol. 76, pp. 817-823, 1981.

[4] D. Hubel and T. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *Journal of Physiol.*, vol. 160, 1962.

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

[6] S. Mallat and Z. Zhang "Matching Pursuit with Time-Frequency Dictionaries", *IEEE Trans. on Signal Processing*, Dec. 1993.

[7] Y. C. Pati R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal Matching Pursuit: Recursive Function Approximation with Applications to Wavelet Decomposition," *Proceedings of the* 27^{th} *Annual Asilomar Conference on Signals, Systems, and Computers*, Nov. 1993.

[8] D. A. Pollen and S. F. Ronner, "Visual cortical neurons as localized spatial frequency filter," *IEEE Trans. Syst., Man, Cybern.*, vol. 13, Sept. 1983.