Curvelets,
Multiresolution Representation, and Scaling Laws

Emmanuel J. Candès and David L. Donoho

Department of Statistics
Stanford University
Stanford, CA 94305-4065, USA

ABSTRACT

Curvelets provide a new multiresolution representation with several features that set them apart from existing representations such as wavelets, multiwavelets, steerable pyramids, and so on. They are based on an anisotropic notion of scaling. The frame elements exhibit very high direction sensitivity and are highly anisotropic. In this paper we describe these properties and indicate why they may be important for both theory and applications.

Keywords: Edges. Partitioning. Subband Filtering. Local Fourier Transform. Ridge functions. Ridgelets. Multiscale ridgelets. Pyramids.

1. INTRODUCTION

Images have edges. Viewing an image as a mathematical object, one might think of an image as an otherwise smooth function with discontinuities along curves, a description which would not be suitable, however, for all kinds of images. Indeed, images of natural scenes are more than just smooth luminance surfaces separated by step discontinuities; for instance, we do not make any claims about the representation of textures, which are typically unsmooth. Our intention is rather to describe a class of images where edges are clearly the dominating features, e.g. cartoons or geometric images.

1.1. Our Viewpoint: Harmonic Analysis

The curvelet construction[1,2] was originally developed for providing efficient representations of smooth objects with discontinuities along curves; the underlying motivation being to apply this construction to classical image processing problems such as:

- *Data compression.* Compression of digitally acquired image data.
- *Image restoration, image reconstruction or edge-preserving regularization.* Enhancement (noise removal) of digital images possibly obtained via indirect measurements as in tomography.[3] One of the main challenges here is to develop smoothing or reconstruction techniques that would smooth out the 'flat' part of the image without blurring the edges.

Our viewpoint is that of modern harmonic analysis whose aim is to develop new representation systems. That is, one is searching for new collection of templates or "elementary forms" that will serve for both the analysis and the synthesis of an object under study. The most classical example is, of course, that of

Further author information: Send correspondence to emmanuel@stat.stanford.edu

Fourier series where an object, in $L_2[0,1]^2$ is expanded out in a superposition of sinusoids according to the rule

$$f(x) = \sum_{n \in \mathbb{Z}^2} c_n(f) e^{i2\pi n \cdot x}, \quad c_n(f) = \int f(x) e^{-i2\pi n \cdot x}\, dx,$$

where n indexes the doubles $(n_1, n_2) \in \mathbb{Z}^2$. The Fourier transform offers an alternate viewpoint to the spatial description of the signal and opens up the possibility of developing new schemes based on the processing of Fourier coefficients. This approach is powerful and in image processing, the two-dimensional Fourier transform and analogs are at the core of a countless number of algorithms. For instance, blocking a digital image into disjoint arrays of 8 by 8 pixels and coding the coefficients obtained after applying the Discrete Cosine Transform to those arrays used to be the of basis of JPEG, the former image compression standard.

In the last decades, many representation systems have been introduced as alternatives to the 'classical' Fourier representation; among those, the Gabor system (time-frequency viewpoint) and wavelets (time-scale viewpoint) probably occupy a prominent place. Wavelets have had a wide impact, both in theory and in practice, and especially in the our areas of preoccupation, namely, data compression and signal restoration. The shrinking of wavelet coefficients, originating in the work of Donoho and Johnstone,[4] proved to be a very powerful tool for statistical estimation, from both a theoretical and a practical standpoint. Similarly, wavelet based-coders have found wide applications in various data compression applications and have been included in JPEG-2000, the newly developed still picture compression standard.

1.2. Our Grail

Within this framework, we aim at finding a representation that is 'optimal' for representing objects with discontinuities along curves. We need, however, to explain what we mean by 'optimal' representation.

First, the concept of representation is classical in harmonic analysis and consists of finding a system (f_μ) and a rule such that any function can be represented as follows:

$$f = \sum_\mu c_\mu(f) f_\mu; \tag{1}$$

classically, one would require that

1. the coefficients c_μ are bounded linear functionals of f, and
2. the decomposition is stable in the sense that one has a quasi-Parseval relation

$$\sum_\mu |c_\mu(f)|^2 \sim \|f\|_{L_2}^2.$$

Second, by 'optimal' representation for a class of objects, we mean that the coefficients of the objects in question are as sparse as possible. In nontechnical terms, this means that, when reordered by decreasing amplitude, the coefficient sequence decays as rapidly as possible.

1.3. Three Anomalies

About twenty years ago, Burt and Adelson introduced the Laplacian pyramid, a 'revolutionary' idea which triggered a series of completely new algorithms in numerical image processing. Canonical pyramid ideas are perhaps known today under the name of wavelets, although a few of these ideas were introduced before the name 'wavelet' was coined. In that sense, we follow Meyer[5] in viewing wavelet theory as a unifying mathematical language for describing a set of connected ideas that arose in different areas.

Although we recognize the importance and the wide impact of these ideas, we feel, nevertheless, that these canonical pyramid ideas are part of a larger picture; we intend to make clear how much larger the question is than just the portion which wavelets represent. In addition, we also believe that a lot of claims regarding the applicability of wavelets to image processing problems such as those mentioned above have been perhaps overstated.

1. *Inefficient Representations.* From a theoretical viewpoint, wavelet series are not optimal for representing objects with discontinuities along curves. Although the next section develops a heuristic argument of the reason, we do not dwell further on this issue and simply refer the reader to existing literature on the subject, see[6] for example.

2. *Limitations of Existing Pyramid Schemes.* Traditional pyramids have only a fixed number of directional elements, independent of scale.

3. *Limitations of Existing Scaling Concepts.* Traditional pyramids do not have highly anisotropic elements.

The point of this paper is to show that it is possible to build a pyramid that is very different that (2)-(3). In fact, the paper will exhibit a pyramid with scaling properties that are very different.

1.4. Outline and Claims

In a recent article,[1] the authors introduced a new system, namely, the curvelet frame which turns out to be very different from existing ideas. Curvelets provide a new multiresolution representation that set them apart from existing representations such as wavelets. The scope of this paper is to review the construction of curvelets and to detail some of their properties. More precisely, our intention is to develop the following claims:

- *Near-optimal representation.* Curvelets provide optimally sparse representation of otherwise smooth objects.
- *Different from existing ideas.* The curvelet transform corresponds to a new way of processing data, unlike any in practice.
- Curvelets have an interesting structure vis-a-vis
  - Scale Space,
  - Pyramids,
  - Non-quadratic image smoothing.

A series of recent articles[1,3,2] addressed the first of these three points and, therefore, we shall here be especially concerned with the other two.

2. LIMITATIONS OF EXISTING IMAGE REPRESENTATIONS

In the introduction, we pointed out that wavelet algorithms only address a portion of the range of pyramids one can dream of. The scope of this section is roughly to describe the specificity and limitations of wavelet-like algorithms.

2.1. Inefficiency of Existing Image Representations

Let us describe a simple heuristic argument that may help explaining the reason why wavelets fail at efficiently representing objects with edges. As we will see, the argument will appear to be rather helpful for motivating the basic geometric features of the curvelet frame.

Suppose that an image f is decomposed into subbands:

$$f_j = D_j(f), \quad j \in \mathbb{Z}$$

where D_j is a filter extracting components of f in the corona of frequencies $|\xi| \in [2^j, 2^{j+1}]$. (The idea of decomposing a signal into dyadic subbands goes back a long way, probably to Littlewood Paley, in the thirties.) The outcome f_j of the passband filter is an image with the following specificity: the energy of f_j is mostly concentrated along the edges of the image f. This is well-known to anyone with experience –even limited– in image processing. For instance, if f is the indicator function of a set B with a smooth boundary, the support of f_j is essentially supported in a region of width proportional to 2^{-j} around the edge ∂B. The connection between edges and the localization of each subband is, of course, not new. In their pioneering work, Marr and Hildreth[7] proposed a multiscale algorithm for the detection of edges, that is to say, of intensity changes; because intensity changes occur at different scales, the algorithm tracks, at each scale j, the curves along which the intensity of the subband f_j crosses zero.

In a wavelet representation scenario, those wavelets at scale j will, roughly speaking, serve to reconstruct the j-th subband f_j. The inadequacy of wavelets for representing an image with an edge now becomes apparent. Two-dimensional wavelets are oscillatory isotropic bumps; at scale j, they are essentially supported inside a disc of radius proportional to 2^{-j}. The subband f_j to be reconstructed, however, is a very elongated band of width 2^{-j} and, instead of isotropic bumps, one would obviously prefer using a system with elements having the following features

- directional sensitivity, and
- high anisotropy.

2.2. Limitations of Existing Pyramid Schemes

The following list of algorithms seems to be representative of the variety of pyramid ideas that were developed in a time span going from the early eighties to the present:

- The Laplacian Pyramid of Burt and Adelson,[8]
- The Wavelet Orthogonal Pyramid of Mallat and Meyer,[9]
- Steerable Pyramids of Adelson, Heeger and Simoncelli,[10]
- Multiwavelets of Alpert, Beylkin, Coifman and Rokhlin.[11]

All of these algorithms share the following common features:

- they have elements at dyadic scales and dyadic locations, and
- they have a *fixed number* of elements at each scale/location.

We say use here the word 'limitation' because it seems plausible to envision pyramids with different properties. e.g. with a number of elements at each scale/location that is scale dependent.

2.3. Limitations of Existing Scaling Concepts

In the literature of wavelet analysis and image analysis, one must frankly admits that there exists a prevailing notion of scaling. To fix ideas, let us consider the case of the orthonormal wavelet transform. A wavelet coefficient is calculated as follows:

$$\int f(x_1, x_2) 2^j \psi(2^j x_1 - k_1, 2^j x_2 - k_2)\, dx_1 dx_2 = \int T_Q f(x_1, x_2) \psi(x_1, x_2)\, dx_1 dx_2$$

where T_Q is the transport operator

$$T_Q f = 2^j f(2^j x_1 + k_1, 2^j x_2 + k_2).$$

Hence, a coefficient is obtained by evaluating the inner product between a dilated and translated version of f and of a single function ψ (or of a fixed number in the case of multiwavelets). The dilation is isotropic and corresponds to the traditional notion of scaling in current use:

$$f_a(x_1, x_2) = f(ax_1, ax_2), \qquad a > 0.$$

However, curves exhibit a very different kind of scaling. For instance let f be the binary function defined by

$$f(x_1, x_2) = 1_{\{y \geq x^2\}}.$$

Then the natural scaling law that applies (at the origin) to this object is of the form

$$f_a(x_1, x_2) = f(a \cdot x_1, a^2 x_2),$$

which is known in Harmonic Analysis as a *Parabolic Scaling.*

This last example illustrates the scaling that operates at the origin. In general, the rescaling of a curve must, of course, be locally adaptive. Suppose that we wish to rescale the curve in a neighborhood of a point x_0 on an arbitrary C^2 curve. Then a dilation by a factor of $1/2$ in the direction of the tangent at x_0 must be accompanied with a dilation by a factor of $1/4$ in the normal direction.

To summarize this section, curves exhibit a kind of scaling which is

- anisotropic, and
- locally adaptive.

3. CONSTRUCTION OF CURVELETS

Curvelets provide optimally sparse representation of otherwise smooth objects. It is beyond the scope of this paper to give a detailed description of this new system, see[1] for a complete exposition. The curvelet transform is based on combining several ideas:

- *Ridgelets*, a method of analysis suitable for objects with discontinuities across straight lines.[6]
- *Multiscale Ridgelets*, a pyramid of windowed ridgelets, renormalized and transported to a wide range of scales and locations.
- *Bandpass Filtering*, a method of separating an object out into a series of disjoint scales.

Given the space limitation, we can only give a very coarse description of the transform and refer the reader to other articles[1,2] for a more detailed account.

3.1. Ridgelets

The theory of ridgelets was developed in the Ph.D. Thesis of Emmanuel Candès (1998). In that work, Candès showed that one could develop a system of analysis based on ridge functions

$$\psi_{a,b,\theta}(x_1, x_2) = a^{-1/2}\psi((x_1 \cos(\theta) + x_2 \sin(\theta) - b)/a). \tag{2}$$

He introduced a continuous ridgelet transform $R_f(a, b, \theta) = \langle \psi_{a,b,\theta}(x), f \rangle$ with a reproducing formula and a Parseval relation. He showed how to construct frames, giving stable series expansions in terms of a special discrete collection of ridge functions. The approach was general, and gave ridgelet frames for functions in $L^2[0,1]^d$ in all dimensions $d \geq 2$ – for further developments, see.[12,6]

Donoho[13] showed that in two dimensions, by heeding the sampling pattern underlying the ridgelet frame, one could develop an orthonormal set for $L^2(\mathbb{R}^2)$ having the same applications as the original ridgelets. The ortho ridgelets are indexed using $\lambda = (j, k, i, \ell, \epsilon)$, where j indexes the ridge scale, k the ridge location, i the angular scale, and ℓ the angular location; ϵ is a gender token. Roughly speaking, the ortho ridgelets look like pieces of ridgelets (2) which are windowed to lie in discs of radius about 2^i; $\theta_{i,\ell} = \ell/2^i$ is roughly the orientation parameter, and 2^{-j} is roughly the thickness.

A formula for ortho ridgelets can be given in the frequency domain

$$\hat{\rho}_\lambda(\xi) = |\xi|^{-\frac{1}{2}}(\hat{\psi}_{j,k}(|\xi|)w^{\varepsilon}_{i,\ell}(\theta) + \hat{\psi}_{j,k}(-|\xi|)w^{\varepsilon}_{i,\ell}(\theta + \pi))/2 \ . \tag{3}$$

Here the $\psi_{j,k}$ are Meyer wavelets for $\mathbb{R}$,[14,15] $w^{\epsilon}_{i,\ell}$ are periodic wavelets for $[-\pi, \pi)$, indices run as follows: $j, k \in \mathbb{Z}$, $\ell = 0, \ldots, 2^{i-1} - 1$; $i \geq i_0$, and, if $\epsilon = 0$, $i = \max(i_0, j)$, while if $\epsilon = 1$, $i \geq \max(i_0, j)$. Notice the restrictions on the range of i, ℓ. Let Λ denote the set of all such indices λ.

In the space domain, the ridgelet ρ_λ is essentially located near the line $x_1 \cos\theta_{i,\ell} + x_2 \sin\theta_{i,\ell} - k2^{-j} = 0$. Hence, we will say that ρ_λ is oriented in the codirection $\theta_{i,\ell} = \ell/2^i$ and the parameter $k/2^j$ will refer to its location. It is useful to think of an ortho ridgelet as an object which have a “length” of about 1 and a “width” which is 2^{-j} (the width can be arbitrarily fine).

Another important property is the fact that, roughly speaking, the ridgelet ρ_λ for $i > j$ is supported outside the unit square $[0,1]^2$. For a given scale j, the cardinality of the ridgelets supported near $[0,1]^2$ is, therefore, about 2^{2j}:

- There are about 2^j orientations corresponding to codirection $\theta_{j,\ell} = \ell/2^j$, and
- There are about 2^j locations per direction since we want $k/2^j = O(1)$.

3.2. Multiscale Ridgelets

Let Q denote a dyadic square $Q = [k_1/2^s, (k_1+1)/2^s) \times [k_2/2^s, (k_2+1)/2^s)$ and let $\mathcal{Q}$ be the collection of all such dyadic squares. The notation $\mathcal{Q}_s$ will correspond to all dyadic squares of scale s. Let w_Q be a window centered near Q, obtained after dilation and translation of a single w, such that the w_Q^2's, $Q \in \mathcal{Q}_s$, make up a partition of unity. We define multiscale ridgelets by $\{\rho_{Q,\lambda} : s \geq s_0, Q \in \mathcal{Q}_s, \lambda \in \Lambda\}$

$$\rho_{Q,\lambda} = w_Q \, T_Q \rho_\lambda,$$

where

$$T_Q f = 2^s f(2^s x_1 - k_1, 2^s x_2 - k_2).$$

The multiscale ridgelet system renormalizes and transports the ridgelet basis, so that one has a system of elements at all lengths and all finer widths.

3.3. Subband Filtering

The last component required is a bank of filters $(P_0 f, \Delta_1 f, \Delta_2 f, \dots)$ with the property that the passband filter Δ_s is concentrated near frequencies $[2^{2s}, 2^{2s+2}]$, e.g.

$$\Delta_s = \Psi_{2s} * f, \quad \widehat{\Psi_{2s}}(\xi) = \widehat{\Psi}(2^{-2s}\xi).$$

3.4. The Curvelet Transform

The curvelet decomposition can be stated in the following form:

- *Subband Decomposition.* The object f is filtered into subbands:

$$f \mapsto (P_0 f, \Delta_1 f, \Delta_2 f, \dots).$$

- *Smooth Partitioning.* Each subband is smoothly windowed into "squares" of an appropriate scale:

$$\Delta_s f \mapsto (w_Q \Delta_s f)_{Q \in \mathcal{Q}_s}.$$

- *Renormalization.* Each resulting square is renormalized to unit scale

$$g_Q = (T_Q)^{-1}(w_Q \Delta_s f), \qquad Q \in \mathcal{Q}_s. \tag{4}$$

- *Ridgelet Analysis.* Each square is analyzed in the orthonormal ridgelet system.

$$\alpha_\mu = \langle g_Q, \rho_\lambda \rangle, \qquad \mu = (Q, \lambda).$$

Figure 1 gives an overview of the organization of the curvelet transform.

Curvelets are then given by

$$\gamma_\mu = \Delta_s \rho_{\lambda,Q}, \qquad \mu = (\lambda \in \Lambda, Q \in \mathcal{Q}_s). \tag{5}$$

It follows from both the definition of the ridgelets (3) and of the multiscale ridgelets that $\hat{\rho}_{Q,\lambda}$ is supported near the corona $[2^{j+s}, 2^{j+s+1}]$. Hence, the only multiscale ridgelets that survive to the passband filtering are those λs such that the scale j satisfies

$$j + s \approx 2s. \tag{6}$$

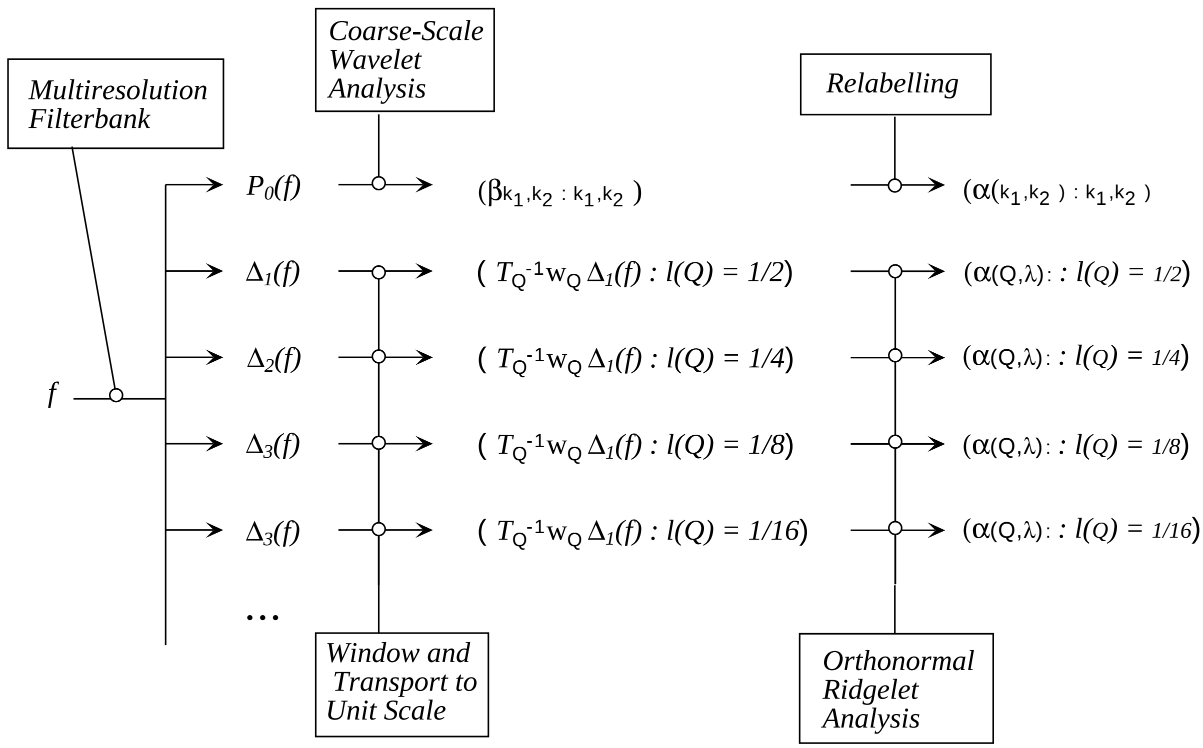

Figure 1. Overview of Organization of the Curvelet Transform.

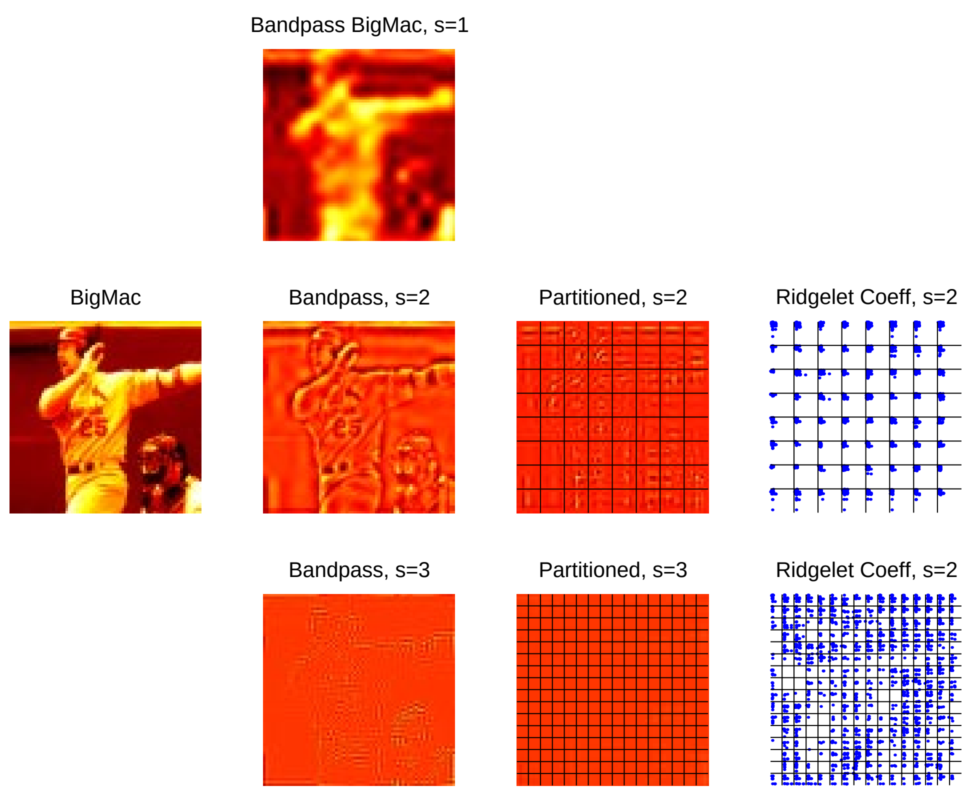

Figure 2. BigMac Image, and stages of Curvelet Analysis.

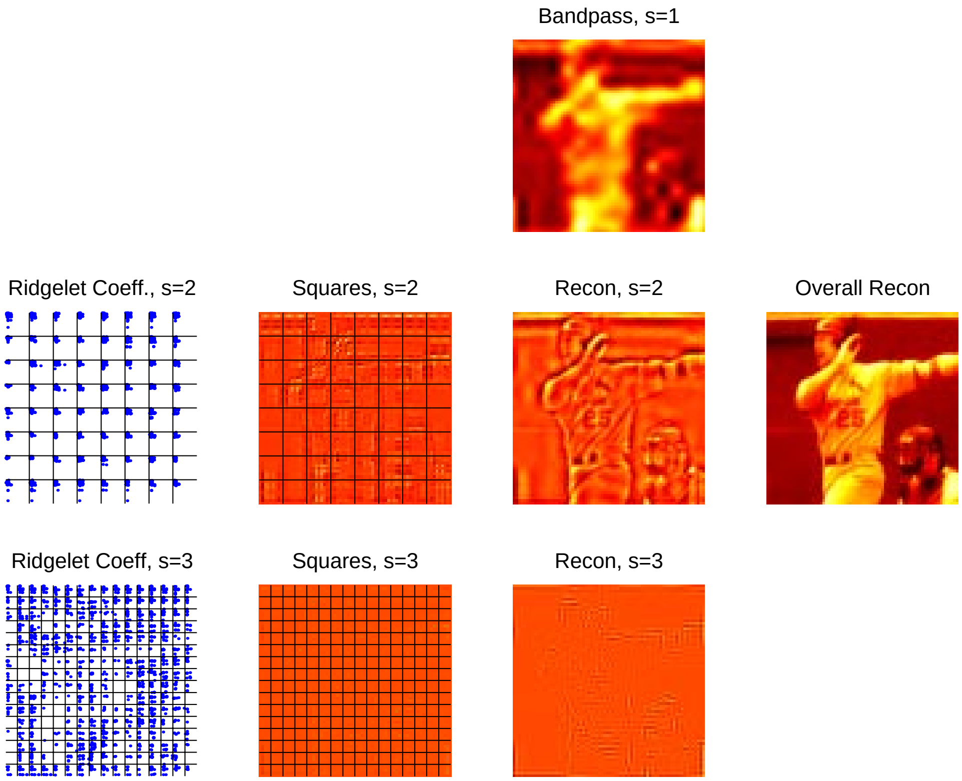

Figure 3. Curvelet Coefficients, and stages of Curvelet Synthesis.

This observation is key for understanding the main properties of the curvelet elements which will be detailed in the next section.

Figure 2 shows how the curvelet transform operates on a digital image; the size of the image is 256 by 256 pixels. The three stages of the curvelet transform are clearly represented.

Figure 3 shows the different stages of the reconstruction of an image from the datum of its curvelet coefficients. The reconstruction is obtained by inverting the three stages of the forward transform, namely,

1. the decomposition into subbands,
2. the spatial localization of each subband,
3. the ridgelet transform of each block, localized in space and frequency.

4. PROPERTIES OF CURVELETS

Curvelets make up a tight frame of $L_2(\mathbb{R}^2)$[1,2]; we have a Parseval relation

$$\sum_{\mu} |\langle f, \gamma_\mu \rangle|^2 = \|f\|_2^2, \tag{7}$$

and a reconstruction principle

$$f = \sum_{\mu} \langle f, \gamma_\mu \rangle \gamma_\mu, \tag{8}$$

with an equality holding in an L_2 sense. Both equalities (7) and (8) are of considerable practical relevance: one can expand out an object into a curvelet series in a very concrete and stable fashion. Although the Parseval relation together with the decomposition formula are reminiscent of an orthonormal decomposition, we would like to emphasize that the curvelet system is redundant, and therefore, not orthonormal.

In addition to the tight frame property, the curvelet transform exhibit an interesting structure that sets it apart from existing image representations:

- The curvelet transform exhibits a new kind of pyramid structure.
- Curvelet frame elements exhibit new scaling laws.
- Curvelets provide an efficient representation of images with edges.

4.1. Pyramid Structure

Figure 2 highlighted the three stages of the curvelet transform. In essence, the first two stages, namely, the decomposition into subbands and the spatial localization of each subband are classical although the coronization $[2^{2s}, 2^{2s+2}]$ associated with the bandpass filters is *nonstandard.* In a classical wavelet pyramid, the coronization would, instead, be of the form $[2^s, 2^{s+1}]$. Bearing this important difference in mind, the curvelet transform might be interpreted as a sequence of two pyramids:

- A first pyramid, indexed by Q whose range is recalled to be the set of all dyadic squares, which localizes the image both in space and frequency;
- A second pyramid, namely, the ridgelet pyramid which analyzes each renormalized block of image data that obey spatial and frequency localization properties (they were denoted by g_Q in the previous section) using directional and anisotropic elements.

First, although the interpretation of the scale may –at this stage– appear somewhat ambiguous, the transform displays the classical notions of

- dyadic scale, and
- dyadic location.

Second, the ridgelet transform adds an original feature to the pyramid: the latter transform introduces two new ingredients, namely,

- direction, and
- microlocation.

Our terminology requires some explanation. There are two levels of location associated with the curvelet transform: a coarse level corresponding to the dyadic square Q which gives an approximate location of a curvelet; and a finer one which gives the location of the curvelet relative to this dyadic square, a piece of information given by the parameter $k/2^j$. The word 'microlocation' refers to this finer level.

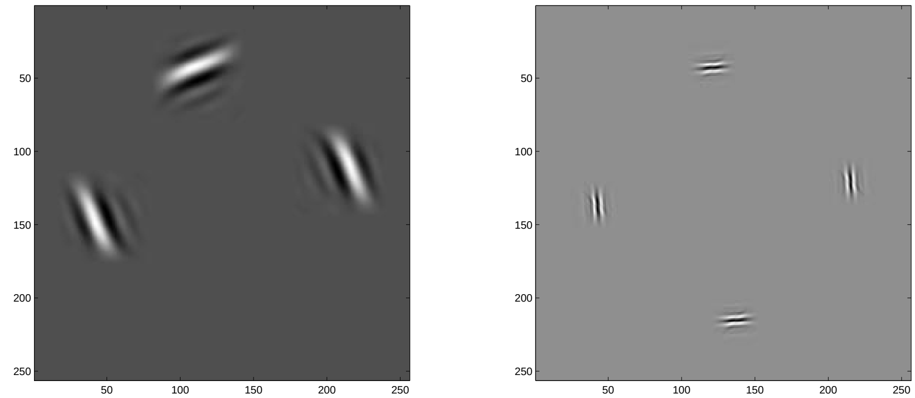

Figure 4. Representation of a few curvelets at two different scales. The left panel corresponds to a scale $s = 2$, while the right panel corresponds to a scale $s = 3$.

4.2. Scaling Laws

In the curvelet pyramid, let us define the scale as being the sidelength of the dyadic square Q indexing a curvelet element. With this definition and (5), we have that the scale of a curvelet is roughly equal to its length. In other words,

$$length(\gamma_\mu) \approx 2^{-s}. \tag{9}$$

The following anisotropy scaling relation is key to the construction; curvelets have support obeying the scaling law

$$width \approx length^2 \tag{10}$$

Hence, the anisotropy is increasing with decreasing scales according to a quadratic power law. This property seems to be new in the literature of computational harmonic analysis as we are not aware of any other system obeying the same scaling relation. Figure 4 pictures a few curvelets at two different scales.

The quadratic scaling relation follows from the earlier observation (6). A curvelet is supported near the dyadic square Q and thus, its length is effectively that of Q, i.e. 2^{-s}. Equation (6) now gives us its approximate width the width of $\rho_{Q,\lambda}$ is about $2^{-s-j} \approx 2^{-2s}$, which justifies (10).

This same principle, namely, (6) gives two additional scaling relations. Because $j \approx s$, it follows from the definition of a curvelet coefficient

$$\alpha_\mu = \langle g_Q, \rho_\lambda \rangle$$

(where g_Q is given by (4)) that

- the number of directions is about proportional to the inverse of the scale, and
- the number of microlocations is about proportional to the inverse of the scale.

Another way, perhaps more insightful, to explain the scaling relations operating within the curvelet transform is to examine the transition from one scale to the next finer scale, i.e. from 2^{-s} to 2^{-s-1}. Each refinement of scale

- doubles the spatial resolution; that is, the size of the dyadic squares in the pyramid is reduced by a factor of two (much like wavelet pyramids).
- doubles the angular resolution; that is, the number of directions of the anisotropic analyzing elements is increased by a factor of two.

This combination is new and of independent interest.

5. CONCLUSION

In this paper we showed that new kinds of pyramids exist with properties, unlike any in current use. We exhibited one such new pyramid, the curvelet pyramid which derives from ridgelet analysis. Beyond the novelty of the curvelet transform, it is important to keep in mind that the construction was tailored for providing efficient representations of images with edges. In this direction, very precise results about the optimality of curvelets have been proved in another article.[1] Short of these results, our new construction would have been perhaps elegant, but not fully justified.

Finally, because of space limitations, we were unable to report on numerical experiments but hope to do so in a later paper. We invite the reader to visit our webpage at
`http://www-stat.stanford.edu/ mduncan/curvelet.site/` for some early numerical results.

ACKNOWLEDGMENTS

This research was supported by National Science Foundation grants DMS 98–72890 (KDI) and DMS 95–05151; and by AFOSR MURI-95-P49620-96-1-0028.

REFERENCES

1. E. J. Candès and D. L. Donoho, "*Curvelets*." Manuscript. `http://www-stat.stanford.edu/~donoho/Reports/1998/curvelets.zip`, 1999.
2. E. J. Candès and D. L. Donoho, "Curvelets – a surprisingly effective nonadaptive representation for objects with edges," in To appear *Curves and Surfaces*, L. L. S. et al., ed., Nashville, TN, (Vanderbilt University Press), 1999.
3. E. J. Candès and D. L. Donoho, "Recovering edges in ill-posed inverse problems: Optimality of curvelet frames," tech. rep., Department of Statistics, Stanford University, 2000.
4. D. L. Donoho and I. M. Johnstone, "Minimax estimation via wavelet shrinkage," *Ann. Statist.* , 1998.
5. Y. Meyer, *Wavelets: Algorithms and Applications*, SIAM, Philadelphia, 1993.
6. E. J. Candes and D. L. Donoho, "Ridgelets: the Key to Higher-dimensional Intermittency?," *Phil. Trans. R. Soc. Lond. A.* , 1999.
7. D. Marr and E. Hildreth, "Theory of edge detection," *Proc. Royal Society of London B* **207**, pp. 187–217, 1980.
8. P. Burt and E. Adelson, "The laplacian pyramid as a compact image code," *IEEE Trans. Comm.* **COM-31**, pp. 532–540, 1983.
9. S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11**(7), pp. 674–693, 1989.
10. E. Simoncelli, W. Freeman, E. Adelson, and D. Heeger, "Shiftable multi-scale transforms [or "what's wrong with orthonormal wavelets"]," *IEEE Trans. Information Theory* , 1992.
11. B. Alpert, G. Beylkin, R. Coifman, and V. Rokhlin, "Wavelet-like bases for the fast solution of second-kind integral equations," *SIAM J. Sci. Comput.* , 1993.
12. E. J. Candès, "Harmonic analysis of neural netwoks," *Applied and Computational Harmonic Analysis* **6**, pp. 197–218, 1999.
13. D. L. Donoho, "*Orthonormal ridgelets and linear singularities*," tech. rep., Department of Statistics, Stanford University, 1998. Submitted for publication.
14. P. G. Lemarié and Y. Meyer, "Ondelettes et bases hilbertiennes," *Rev. Mat. Ibero-americana* , 1986.
15. Y. Meyer, *Wavelets and Operators*, Cambridge University Press, 1992.