Orientational Minimal Redundancy Wavelets: from edge detection to perception

Antonio Turiel

Air Project - INRIA
Domaine de Voluceau BP105
78153 Le Chesnay CEDEX
France

Jean-Pierre Nadal

Laboratoire de Physique Statistique, Ecole Normale Supérieure
24 rue Lhomond
75231 Paris CEDEX 05
France

Néstor Parga *

Departamento de Física Teórica
Universidad Autónoma de Madrid
28049 Cantoblanco, Madrid
Spain

Abstract

Natural images are complex but very structured objects and, in spite of its complexity, the sensory areas in the neocortex in mammals are able to devise learned strategies to encode them efficiently. How is this goal achieved? In this paper, we will discuss the multiscaling approach, which has been recently used to derive a redundancy reducing wavelet basis. This kind of representation can be statistically learned from the data and is optimally adapted for image coding; besides, it presents some remarkable features found in the visual pathway. We will show that the introduction of oriented wavelets is necessary to provide a complete description, which stresses the role of the wavelets as edge detectors.

Key words: Wavelet, multiscale, edge detection, learning

1 Introduction

In recent years there has been much work at the boundary between the modeling of visual systems in mammals and computer vision. On the one hand a better knowledge of natural systems may lead to new image processing algorithms, and on the other hand analysis and modelling of images may help to understand the primary layers in visual systems.

Based on early papers by Barlow (Barlow, 1961), many works have focussed on the use of information theoretic concepts in order to address the question of efficiency of neural coding. Information Theory allows to quantify the statistical regularity of a signal - hence the minimal number of bits necessary to encode an image -, the likeliness of an event, and the statistical dependency between signal (stimulus) and neural code (representation). It has been suggested that epigenetic development of the visual system of mammals would aim at optimizing the neural architecture so that the visual system becomes optimally adapted to the statistical regularities of the environment. One criterion for this adaptation is the minimization of the redundancy (Barlow, 1961): each output unit should be as much as possible statistically independent from any other unit. Another plausible criterion is the maximization of information transfer from the retina to the next cortical layers (Linsker, 1988). Such optimization occurs under constraints, in particular under limited resources (e.g. a given set of neurons and neural connectivity). In fact both criteria are related, as it was shown in (Nadal and Parga, 1994) the code which maximizes information transfer minimizes redundancy, that is, it extracts the independent components of the signal.

Several theoretical studies of the primary visual system have been done, based on the ideas of information transfer and redundancy reduction (see e.g. (Atick and Redlich, 1990, 1992) (Atick, 1992), (van Hateren, 1992), (Bell and Sejnowski, 1997), (Olshausen and Field, 1996, 1997)). When modeling the visual inputs as generated by a Gaussian process, it has been shown that the additional hypothesis of scale invariance leads to a scale invariant, wavevet-like, representation (Li and Atick, 1994). This results from a compromise between scale and translational invariances, which cannot be exactly fulfilled at the same time. Direct statistical analysis of natural images leads also naturally to a multiscale description, see (Field, 1987), (Ruderman, 1994), (Turiel et al., 1997, 1998), (Buccigrossi and Simoncelli, 1999), (Huang and Mumford, 1999).

As images consist of objects, which are two-dimensional regions with regular border, the ratio of the length of the border to the surface that it contains increases with scale. So,

* To whom correspondence should be addressed
Email addresses: Antonio.Turiel@inria.fr (Antonio Turiel),
Jean-Pierre.Nadal@lps.ens.fr (Jean-Pierre Nadal), parga@delta.ft.uam.es (Néstor Parga).

the statistics of the illumination is dominated by the smooth changes taking place inside objects. This part is however somewhat predictable and thus scarcely informative. The statistical regularity of illumination is evidenced by the behaviour of the power spectrum (which corresponds to the Fourier transform of the two-point correlation of illumination). It is well known that it behaves like a power-law in the modulus of the frequency vector, with a power exponent close to -2 (Field, 1987). This is the typical scale-invariant behaviour for piece-wise regular functions.

In the search for an independent representation, a first step is to perform whitening (Atick, 1992): a linear transformation on the input signal which leads to a flat power spectrum (decorrelation of the two-point statistics). Experiments on mammal cortices have shown that in fact some withening-like process takes place as the image is transmitted to the visual cortex (see e.g. (Dan et al., 1996)).

However images are not Gaussian. Indeed contours are still recognizable features of whitened images (B.Barlow, 1994), the variation of the illumination along them being still very regular. Contours, more generally edges, carry the most important information about the image, although they are very scarce. This means that the statistics of images is in fact controlled by those rare events. This far-from-gaussianity character of images has been reported for instance as exponential decays of several probability distributions (Ruderman, 1994). Since linear addition of independently distributed features would typically give rise to a Gaussian behaviour, one also expects that the extraction of independent features will require non-linear processing.

In previous studies, a multifractal analysis of natural images has been performed (Turiel et al., 1997, 1998), (Turiel and Parga, 2000a) and tested on a wide variety of ensembles of natural images (Nevado et al., 2000), (Turiel et al., 2000). It has been shown that an optimal wavelet (Turiel and Parga, 2000b) can be constructed (learned) from a set of images. The resulting wavelet representation achieves both whitening and edge detection. More importantly, the dyadic expansion on this wavelet splits the image in statistically independent components, one per level of resolution. This representation has thus several important features shared by the neural representation in mammals. However the wavelet basis is not complete, which implies some information loss.

In this work we propose a generalization of the wavelet representation introduced in (Turiel and Parga, 2000b): we show that a complete basis can be obtained once orientation detectors are introduced. By means of the generalized oriented multifractal wavelet basis, it is then possible to reconstruct correctly visual signals from their wavelet coefficients.

The paper is organized as follows: in the following Section we review very briefly our previous work. We first describe how a generative model of images containing second order as well higher order scale invariance can be constructed. In the same section we compare

our work with other approaches to the problem. In Section 3 the concept of optimal wavelet, which allows to split the image in independent levels of resolution, is explained and its properties are shown and discussed. In Section 4 the experimental performance of the optimal wavelet is shown, and its lack of completeness is discussed. Section 5 is devoted to illustrate for a simple basis (the Haar basis) how introducing new orientations gives a complete description of the image. Then, in Section 6 the generalization of the optimal wavelet formula to include the role of the orientation is presented and its connection with the previous scheme is shown. Section 7 presents the theory for computing the orientational basis from the data. The experimental facts concerning oriented wavelets are then presented in Section 8. Finally, in Section 9 we discuss the results from the point of view of the empirical performance of the method and of its biological relevance, and we propose future directions of research.

2 Scale invariance in natural scenes

One of the most prominent properties of natural images is scale invariance. Let $P[\{c(\vec{x})\}_{\vec{x}}]$ be the joint probability distribution for the values of a function $c(\vec{x})$ of the image at all the points $\vec{x}$. For images with intensity $I(\vec{x})$ (i.e., graylevel in digitized images) $c(\vec{x})$ is the contrast at the pixel $\vec{x}$ which we take as $c(\vec{x}) \equiv I(\vec{x}) - I_0$. The constant I_0 is chosen such that the average of c over the image vanishes. Scale invariance means that $c(\vec{x})$ is equally distributed as its dilation by a factor r, $c(r\vec{x})$ (see e.g. (Ruderman, 1994)), that is,

$$P[\{c(\vec{x})\}_{\vec{x}}] = P[\{r^{\nu} c(r\vec{x})\}_{\vec{x}}] \tag{1}$$

Scale invariance shows up in the second order statistics as an algebraic behavior of the power spectrum (Field, 1987). However the scale-invariant properties of images are much richer than that. As it was shown in, e.g., (Ruderman, 1994; Turiel et al., 1997, 1998) a complete description of scale invariance requires the analysis of the higher order, non-gaussian statistics.

In natural scenes scale invariance takes the form of multiscaling (or multifractality). Multiscaling happens when images are composed of scale-invariant objects, however transforming differently under changes in scale. Multiscaling can be assessed from the calculation of the order q moments of a random variable $e_r(\vec{x})$ defined at a scale r. Due to scale invariance, those moments can only depend on the scale r as a power law:

$$< {e_r}^q > \sim \ r^{\gamma(q)} \tag{2}$$

The dependence of the self-similarity exponents $\gamma(q)$ on the order q provides knowledge about the underlying multiscaling hierarchy. The simplest manifestation of this symmetry occurs when $\gamma(q)$ depens linearly in the order of the moments (monoscaling). This form of self-similarity is related to the existence of a single underlying scale-invariant object of fractal nature. When the exponents adopt a non-linear dependence, that is, when actual multiscaling occurs, more than one exponent is needed to describe the way the variable changes with the scale; equivalently, images are composed by more than one scale-invariant fractal object. Such images are multifractals. In fact, images have been shown to be always multiscaling (Turiel et al., 1997, 1998; Nevado et al., 2000; Turiel et al., 2000).

There are many reasons why the statistics of edges should be studied carefully. Edges are good candidates for one of the scale-invariant fractal sets in the multiscaling hierarchy, actually it is the most singular of these sets. Edges are also very relevant in vision. A variable $e_r(\vec{x})$ capable to provide multiscaling self-similarity exponents should be devised in such a way that it integrates edges and other potentially scale-invariant features over a scale r. For this reason, a variable (the 'local edge variance") taking account all the changes in contrast over a local area of size r was studied in (Turiel et al., 1997, 1998). Such a variable proved to be multiscaling in experiments performed over several different ensembles of images (Nevado et al., 2000; Turiel et al., 2000).

Some of the main results of that first work have been that: (1) the local edge variances are in fact wavelet coefficients (Turiel and Parga, 2000a), (2) their behavior with the scale does not depend of the particular wavelet chosen and (3) the statistics of these wavelet coefficients can be explained in terms of an infinitely divisible process (Feller, 1966). Another remarkable property found in (Turiel et al., 1997, 1998) is that although the wavelet coefficients at different scales are correlated (Mallat and Zhong, 1992), the marginals of two of these variables at different scales are related through the following statistical relation:

$$e_r \doteq \eta_{r,L}\, e_L \,, \tag{3}$$

where "$\doteq$" means that both sides have the same distribution. The stochastic variable $\eta_{r,L}$ is a multiplicative process. This means that: **a**) it is independent from e_L and **b**) it is an infinitely divisible process: given three scales $L > r' > r$, it satisfies $\eta_{r,L} \doteq \eta_{r,r'}\eta_{r',L}$ (Novikov, 1994). Besides, because of scale invariance, its distribution only depends on the ratio of scales $\frac{L}{r}$.

As was just said, the local edge variance can be easily generalized to wavelet projections, which are experimentally shown to be also multiscaling and give rise to the same hierarchy as the local edge variance (provided the wavelet belongs to an appropriate class; see (Turiel and Parga, 2000a)). This implies, in particular, that the multiplicative process eq. (3)

still holds taking the variables e_r and e_L as the wavelet projections at the scales r and L, respectively. Even more, the multiplicative process is of the same type, that is, it has the same distribution as that of the local edge variance.

Multiscaling, through eq. (3), places strong constrains on what a natural scene is. When it is combined with other two symmetries of natural images -scale and translational invariances- a powerful generative model is obtained (Nevado et al., 2000; Turiel and Parga, 2000b). The resulting model contains *both* second order scaling (the power spectrum follows a power law) and higher order scaling (multiscaling). The model is defined as a discrete (dyadic) wavelet expansion with wavelet coefficients $\alpha_{j\vec{k}}$ (playing the role of the previous e_r) which follow a multiscale stochastic process. More precisely, given a function (wavelet) Ψ the contrast is expressed as

$$c(\vec{x}) \;=\; \sum_{j=0}^{\infty} \sum_{\vec{k}\in(Z_{2^j})^2} \alpha_{j\vec{k}}\, \Psi_{j\vec{k}}(\vec{x}) \;, \tag{4}$$

where $\Psi_{j\vec{k}}(\vec{x}) \;\equiv\; \Psi(2^j\vec{x} - \vec{k})$. The largest scale is fixed as one, and the j-th scale is then 2^{-j}. Assuming that the dispersion of the wavelet is of the same order as the scale, it is possible to distinguish up to 2^j different blocks along each dimension (2^{2j} block in our case, as images are bi-dimensional). The $\alpha_{j\vec{k}}$'s are put in correspondence with the nodes of a tree in such a way that the resolution becomes better as one proceeds from the root to the leaves of the tree. As an extension of eq. (3), the four wavelet coefficients at the scale j are obtained from the one at the previous scale as:

$$\alpha_{j\vec{k}} \;=\; \eta_{j\vec{k}}\, \alpha_{j-1\left[\frac{\vec{k}}{2}\right]} \tag{5}$$

Also here the variables $\eta_{j\vec{k}}$ are independent from the $\alpha_{j-1\left[\frac{\vec{k}}{2}\right]}$ and have the same distribution for all the resolution levels j and spatial locations $\vec{k}$. This distribution is determined from the analysis of multiscaling properties in the image dataset; see (Turiel et al., 1997; Nevado et al., 2000) for further details. Notice that in the generative model the equality of both sides holds exactly, and not only in a distributional sense. The generative model succeeds to incorporate multiscaling behavior and the correct power spectrum.

Not every function Ψ can be used to represent arbitrary signals $c(\vec{x})$ but it should meet some requirements; these imply that the function must reach a compromise between localization and detail detection (i.e., the spatial and frequency dispersions are kept small enough) (for technical details, the reader is encouraged to read the excellent book from I. Daubechies (Daubechies, 1992)). For some particular wavelets Ψ, there exists an associ-

ated dual function $\tilde{\Psi}$ expanding a wavelet basis $\{\tilde{\Psi}_{j\vec{k}}\}$ such that the coefficients $\alpha_{j\vec{k}}$ can be retrieved by simple wavelet projection on $\tilde{\Psi}$, namely:

$$\alpha_{j\vec{k}} \;=\; 2^{2j}\langle\tilde{\Psi}_{j\vec{k}}|c\rangle \tag{6}$$

Here, the symbol $\langle f_1|f_2\rangle$ means the inner product of the functions f_1 and f_2,

$$\langle f_1|f_2\rangle \;\equiv\; \int d\vec{x}\; f_1(\vec{x})\; f_2(\vec{x}) \tag{7}$$

and the constant 2^{2j} in eq. (6) appears as a consequence of the normalization chosen in the definition of $\Psi_{j\vec{k}}$.

A coding scheme based on wavelet expansions can be proposed under the requirement of *efficient coding*. If efficiency is interpreted as a minimum redundancy principle (Barlow, 1961) the wavelet coefficients (obtained under wavelet projection of images) should be as independent as possible. Independency is a stronger requirement than eq. (3) (which holds for a large class of appropriate wavelets). We will require eq. (5) to hold, which is not true for a generic wavelet. The following efficiency criterion was then proposed in (Turiel and Parga, 2000b,c): *the redundancy between wavelet coefficients at different scales should be minimized*. This defines our optimal encoder. Notice that nothing is said about spatial correlations at the same scale, these should be estimated once the optimal wavelet is known and if they are large a minimum redundancy criterion should be applied again.

2.1 Related work

Our approach to natural image modeling differs in many respects from other previous and later approaches to this problem.

The emergence of oriented edge detectors from the statistics of images has been considered before by many authors. We refer in particular to two approaches: one based on the independendent components analysis of natural scences (Bell and Sejnowski, 1997) and the other based on the requirement of the sparseness of the representation (Olshausen and Field, 1996, 1997). In the first of these works the independent components have been found by using the infomax principle for redundancy reduction (Nadal and Parga, 1994) assuming a linear generative model of images. This approach does not consider the role of symmetries of natural scences and much of the insight that can be gained from them is lost. In particular it does not take into account the constraints that multicaling imposes on algorithms to reduce redundancy. A consequence of these constraints is that a factorial

code cannot be obtained by a linear transformation (Turiel and Parga, 2000b). In fact, a comparison of the edge detectors found in this way with the properties of V1 cells (van Hateren and van der Schaaf, 1998) shows that the predicted filters do not reproduce well the scaling properties of cells. Our work differs from (Olshausen and Field, 1996, 1997) in that in our approach sparseness is not imposed but it appears as a consequence of the sparseness of edges in natural scenes, as it is clear from the kurtotic distributions in (Turiel et al., 1997, 1998). It seems to us that the theoretical analysis done for the moment is still far from being complete, natural images still have basic regularities and more subtle symmetries that have to be discovered and used correctly before the predictions of the model could be trusted.

Following the observation that although the wavelet transform gives an approximately decorrelated signal (Wornell, 1993) the wavelet coefficients are still correlated, the authors of (Buccigrossi and Simoncelli, 1999; Wainwright and Simoncelli, 2000) noticed that the dependencies between pairs of adjacent coefficients could be reduced by taking appropriate ratios of neighboring coefficients. This is in agreement with our finding (Turiel et al., 1997, 1998) of an independent multiplicative process, eqs. (3) and (5), although their empirical observation does not lead to the actual formal structure behind it.

The scale mixtures of gaussians used in a recent paper (Wainwright et al., 2001) are stochastic processes similar to those in eq. (3). There are however substantial differences in both the approach to the problem and the nature of the stochastic processes themselves. Our basic guidelines for understanding non-gaussian image statistics have been the analysis of their scaling properties. According to this, we have started in (Turiel et al., 1997) by studying in which way these properties manifest themselves in natural scences. After noticing that they take the form of multiscaling (eq. (2)), we have made extensive analyses to verify that multiscaling is a consequence of the presence of an infinitely divisible process that acts multiplicatively between different scales (as it was briefly reviewed in the previous section). In fact one can argue that multiscaling necessarily implies the existence of such a multiplicative process, so any realistic description of image statistics should be implemented on the basis of an infinitely divisible multiplicative process. Among those processes, the log-Poisson is the simplest one that fits well the data (Turiel et al., 1997, 1998; Turiel and Parga, 2000b). Instead, in (Wainwright et al., 2001) the authors start by proposing a rather wide class of models which however are not compatible with the actual multiscaling properties in images. The most appropriate model would be that of the lognormal family (lognormal distributions are infinitely divisible multiplicative processes), but lognormal statistics introduces uncontrolled, unrealistic divergences to infinity. It could be argued that, according to the Central Limit Theorem, any infinitely divisible process approaches a lognormal distribution when the change in scale is large enough; however the convergence is only guaranteed for the most probable events, the tails being always ill-described. As tails are related to the edges (the kurtotic character

of log-Poisson statistics being at the origin of their sparseness), a lognormal description is far from appropriate for images.

3 The optimal wavelet

The optimal wavelet is determined by requiring the statistical equality in eq. (5) to hold *point-by-point*, that is, the equality is true for any image, resolution and location. This is a very strong statement and in fact it completely determines an unique wavelet Ψ, the optimal wavelet [1] . One can obtain the coefficients $\eta_{j\vec{k}}$ from the exact relation,

$$\eta_{j\vec{k}} = \alpha_{j\vec{k}} / \alpha_{j-1[\frac{\vec{k}}{2}]} \tag{8}$$

These variables $\eta_{j\vec{k}}$ provide a representation in which each level of resolution is independent of the other - $\eta_{j\vec{k}}$ and $\eta_{j'\vec{k}'}$ are statistically independent for $j \neq j'$ (Turiel and Parga, 2000b). For this reason, we will refer to this wavelet representation as the optimal one.

Using the representation provided by the optimal wavelet Ψ, it is possible to design a multilayer network architecture to extract the independent components of the signal. An example architecture is shown in Figure 1. It consists of three layers, two of them with linear transfer functions and the other with a non-linear, e.g. logarithmic transfer. Such non-linearity has been suggested to fit the response of simple cells (Maffei and Fiorentini, 1973). As we will see, this representation has other remarkable features, as learning capability and independence of the power spectrum. It is somewhat ideal and maybe the visual system does not exactly mimic this type of architecture (it could, for instance, introduce the nonlinearity predicted by multiscaling in differente ways such as divisive normalization (Heeger, 1992; Carandini et al., 1997)), but as this representation is determined by the statistical features of images it is by itself a property of images, so it constitutes the ideal goal for independent feature extraction (hence for optimal coding).

Let us first review the main results in (Turiel and Parga, 2000b). The optimal wavelet is completely defined under the assumption of independency for the coefficients $\eta_{j\vec{k}}$. It can be obtained from an ensemble of images by means of the average of the contrast, $C(\vec{x})$, according to the following relation between the Fourier transforms of the functions Ψ and C:

$$\hat{\Psi}(\vec{f}) = \frac{1}{\overline{|\alpha_{0\vec{0}}|}} \left[\hat{C}(\vec{f}) - \frac{\overline{|\eta|}}{4} \frac{\Lambda(\vec{f})}{\Lambda(\frac{\vec{f}}{2})} \hat{C}(\frac{\vec{f}}{2}) \right] \tag{9}$$

[1] Henceforth, Ψ will always represent the optimal wavelet

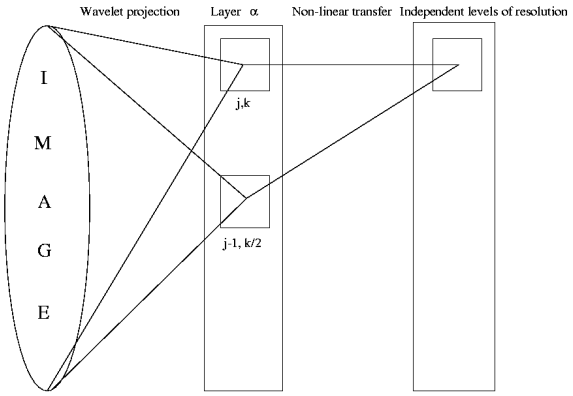

Fig. 1. A possible neural architecture to extract independent features as predicted by scale invariance of natural scenes. The image first stimulates the photoreceptor layer and it is then projected forward by the optimal wavelet to activate the "layer α" cells. However their activity is not necessarily given by this linear transformation since inhibitory interactions between the cells can implement the non-linearity predicted by multiscaling. This can be done, e.g., in the form of divisive normalization (Heeger, 1992; Carandini et al., 1997). Alternatively, a logarithmic transfer could give independent responses of cells coding for features at different scales on a third layer. Only this latter case is shown here.

where $\Lambda(\vec{f}) = (1-e^{-2\pi i f_1})(1-e^{-2\pi i f_2})$, and $\overline{|\eta|}$ is the average of $|\eta|$, which in our context is fixed to $\frac{1}{2}$ for any multifractal structure due to traslational invariance (Turiel and Parga, 2000a).

This expression completely determines the wavelet up to a constant; it is evident that if Ψ is a valid wavelet for performing a dyadic expansion, $a\Psi$ can also be used (for any $a \in \Re$, $a \neq 0$), the coefficients $\alpha_{j\vec{k}}$ being resized by $1/a$ (which leaves unaltered the coeffcients $\eta_{j\vec{k}}$). The undetermination in the constant is resolved by requiring Ψ to be normalized in norm 2, that is,

$$\langle \Psi | \Psi \rangle \;=\; \int d\vec{x}\; \Psi^2(\vec{x}) \;=\; 1 \tag{10}$$

There is still an undetermination in the sign, which we fix conventionally so that the wavelet is always decreasing along the vertical axis (see Figure 2).

In (Turiel and Parga, 2000b) the optimal wavelet was evaluated from an ensemble of 200 images. Once it is obtained, the hypothesis of independence between scales for the variables η's must be self-consistently verified. In (Turiel and Parga, 2000b) the correlation coefficients between $|\eta|$'s at consecutive scales were computed; the values obtained (smaller than 10^{-2} for scales $j > 2$) confirmed the weak dependency between those variables. Better tests of independence are given by the mutual information (Cover and Thomas, 1991).

In (Turiel and Parga, 2000c) it was shown that the mutual information between $|\eta|$'s at consecutive scales is negligible (smaller than 10^{-4} bits for scales beyond $j = 3$, compared to a maximum of 10 bits).

3.1 Properties of the optimal wavelet

As it was just discussed, the wavelet obtained by means of eq. (9) has been empirically shown to lead to $\eta_{j\vec{k}}$ which are independently distributed at each resolution level j (Turiel and Parga, 2000b), self-consistently validating the derivation. It is also remarkable that the wavelet is obtained by on-line learning, as it varies linearly with the average of the contrast over the recorded images (C). Besides, according to eq. (9) the learning capability is cumulative, as the average of the wavelets learnt over two subensembles equals to the wavelet which would be learnt from the joint ensemble. This fact increases the biological plausibility, as it is an indication that the wavelet can be learnt, and its knowledge can be improved in time.

In addition to those two properties, which are consequences of the theoretical scheme, the optimal wavelet we have experimentally obtained possesses other relevant features, which are not necessary for theoretical consistency but which reveal deep properties of natural images. For details on the experimental procedure, see Section 4.

- **The wavelet Ψ is orthogonal:**
  With a small error, the wavelet basis $\{\Psi_{j\vec{k}}\}$ verifies the orthogonality condition,

$$\langle \Psi_{j\vec{k}} | \Psi_{j'\vec{k}'} \rangle \;=\; 2^{-2j} \delta_{jj'} \delta_{\vec{k}\vec{k}'} \tag{11}$$

  This property is very useful from the computational point of view, as it allows to retrieve the coefficients by a simple projection of the signal over the wavelet:

$$\alpha_{j\vec{k}} = 2^{2j} \langle \Psi_{j\vec{k}} | c \rangle \tag{12}$$

  Also, the basis $\{\Psi_{j\vec{k}}\}_{j\vec{k}}$ is an orthogonal basis, so the energy of the image equals the energy of its coefficients

$$\langle c^2 \rangle \;=\; \int d\vec{x}\; c^2(\vec{x}) \;=\; \sum_{j\vec{k}} 2^{-2j} \alpha^2_{j\vec{k}}$$

  The observed empirical discrepancy between the two sides of this equation (see Section 4) is an evidence of the necessity for orientational wavelets (Section 6).
- **The wavelet Ψ acts like an edge detector:**

As it can be observed in Figure 2, the wavelet undergoes a strong transition along the central horizontal line. The coefficients $\alpha_{j\vec{k}}$ obtained by projecting over this wavelet will be greater over the areas of the image where a strong change takes place. On the contrary, due to the opposite signs of the wavelet above and below the horizontal line, the projection will tend to vanish over areas of smooth variation.

This edge-like behaviour of the wavelet is made more evident when the image is regenerated from the coefficients in application of eq. (4), as can be observed in Figures 5, 9 and 10. The image is progressively rebuilt by the addition of edge-like contributions.

- **The modulus of the Fourier transform of Ψ is $1/f$:**

  This fact is strongly connected with very well-known properties of the power spectrum. The power spectrum $S(\vec{f})$ is defined as the Fourier transform of the two-point correlation of c, which for translational invariant fields coincides with the average (over the ensemble of images) of the square of the modulus of the Fourier transform of c; namely:

$$S(\vec{f}) \;=\; \langle |\hat{c}|^2(\vec{f})\rangle \tag{13}$$

  where $\hat{c}(\vec{f})$ stands for the Fourier transform of c at the spatial frequency $\vec{f}$. Introducing the dyadic representation of c, eq. (4), in the definition of the power spectrum, eq. (13), we obtain:

$$S(\vec{f}) \;=\; \sum_{j=0}^{\infty} 2^{-2j}\langle\eta^2\rangle^j \; |\hat{\Psi}|^2(2^{-j}\vec{f}) \tag{14}$$

  The behaviour of the power spectrum is known since the early days of the television; it exhibis a power law of the type (Field, 1987):

$$S(\vec{f}) \sim f^{-(2-\beta)} \tag{15}$$

  where β is a small exponent depending on the ensemble of images considered; frequently there is also a weak deviation from isotropy. It is immediate from eq. (14) that a wavelet Ψ such that $|\hat{\Psi}|(\vec{f}) \sim f^{-1}$ leads to the correct power spectrum (the correction exponent β and the weak anisotropy come out from the uneven weighting for the different orientations in the orientational wavelet expansion, see Section 6). On the contrary, for any wavelet such that $|\hat{\Psi}|$ is different from a power law it follows from eq. (14) that

$$S(2\vec{f}) \approx 2^{-2}\langle\eta^2\rangle S(\vec{f}) \tag{16}$$

  According to (Turiel and Parga, 2000a), $\langle\eta^2\rangle = 2^{-(2+\tau_2)}$ and $-1 < \tau_2 < 0$. We thus obtain $S(2\vec{f}) \approx 2^{-4-\tau_2} S(\vec{f})$ and in general $S(a\vec{f}) \approx a^{-4-\tau_2} S(\vec{f})$, that is, $S(\vec{f}) \sim f^{-4-\tau_2}$. Hence any wavelet such that $|\hat{\Psi}| \neq f^{-1}$ would give rise to an incorrect exponent for the power spectrum (a similar derivation to the one presented here can be found in (Field, 1994)).

The scaling property of $|\hat{\Psi}|$ also allows us to establish a link between the dyadic representation and the reconstruction algorithm proposed in (Turiel and del Pozo, 2002). In that paper, the authors show that images can be reconstructed from the values of contrast changes over the borders (which are identified with the Most Singular Manifold in the multifractal structure (Turiel and Parga, 2000a)). The reconstruction formula is essentially a diffusion of the values of the contrast along the edges according to a kernel which behaves, again, as $1/f$ in Fourier space. As it can be seen in figures 5 , 9 and 10, the wavelet expansion works much in the same way: each resized, translated wavelet appearing in the sum eq. (4) is equivalent to a light-spreading edge element of that size and location, weighted with the appropriated coefficient $\alpha_{j\vec{k}}$.

4 Experimental results

4.1 Determination of the wavelet

The wavelet was evaluated from eq. (9). The normalized average $C(\vec{x})$ was computed over an ensemble of 1000 images taken from Hans van Hateren's web database (see (van Hateren and van der Schaaf, 1998) for details on the images). We took the central 1024×1024 patch of the images labelled *imk00004.imc, imk00008.imc, imk00012.imc,..., imk04000.imc* to obtain Ψ with a resolution of 1024×1024 pixels. As in (Turiel and Parga, 2000b), we observed a rough left-right symmetry and top-bottom antisymmetry. Accepting those two properties to hold, we have stressed them, replacing $\Psi(\vec{x})$ by the following average:

$$\bar{\Psi}(x_1, x_2) \;=\; \frac{\Psi(x_1, x_2) + \Psi(-x_1, x_2) - \Psi(x_1, -x_2) - \Psi(-x_1, -x_2)}{4} \tag{17}$$

were x_1 is the horizontal coordinate and x_2 is the vertical one. In this way $\bar{\Psi}(-x_1, x_2) = \bar{\Psi}(x_1, x_2)$ and $\bar{\Psi}(x_1, -x_2) = -\bar{\Psi}(x_1, x_2)$. The resulting function is represented in Figure 2.

The orientationally averaged modulus of the Fourier transform of the wavelet is represented in Figure 3; the correspondence with a $1/f$ law is almost perfect.

We have checked the orthogonality of the wavelet, eq. (11). As we have just a finite matrix of points, we need to provide a numerical algorithm to compute $\Psi(2^j\vec{x})$. We have devised a method which keeps the $1/f$ dependence in Fourier space at every scale, what we think is an essential property of the wavelet. So, we proceed as follows:

- We apply a whitening filter to the wavelet, that is, we multiply it by the modulus of

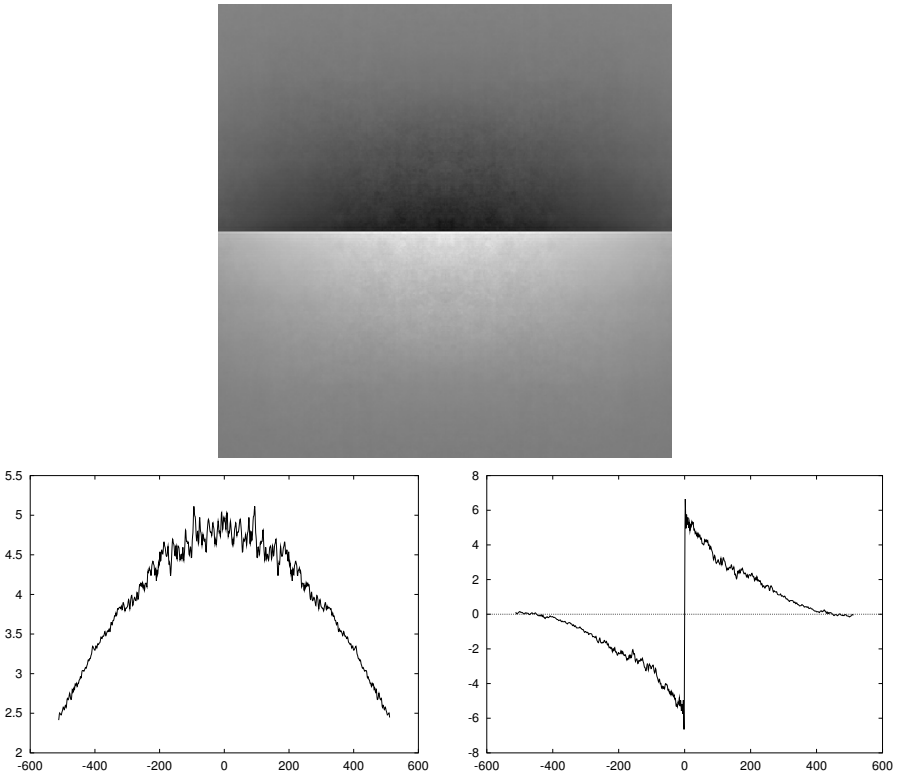

Fig. 2. Top: Gray level representation of the optimal wavelet Ψ
Bottom: Horizontal and vertical cuts of the optimal wavelet.

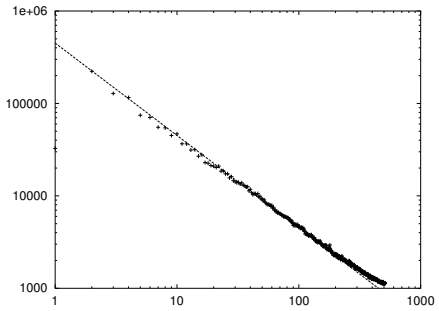

Fig. 3. Orientational average of $|\hat{\Psi}|(\vec{f})$ in log-log scale and best fit with k/f curve, k constant.

the frequency vector in Fourier space:

$$\Psi \to T_1\Psi \quad \text{where} \quad \hat{T_1\Psi}(\vec{f}) \equiv f\hat{\Psi}(\vec{f})$$

- We form a new matrix of $(1024/2^j) \times (1024/2^j)$ blocks, for which each point is the average over the corresponding $2^j \times 2^j$ block of the values of $T_1\Psi(\vec{x})$. We fill the rest of entries of the matrix up to 1024×1024 with zeroes. This is just a way to approximate $T_1\Psi(2^j\vec{x})$.

- We apply to this matrix the inverse of T_1 (denoted as T_{-1}), that is, we multiply its Fourier transform by $1/f$.

The matrix of points so obtained is a discrete, smooth approximation of $\Psi(2^j\vec{x})$ which keeps the $1/f$ dependence in the modulus. Once the algorithm to compute $\Psi(2^j\vec{x})$ is given, we check eq. (11). We define the average error $\in_j$ at the scale j as:

$$\in_j \equiv \sum_{\vec{k}} |\langle\Psi|\Psi_{j\vec{k}}\rangle| \tag{18}$$

that is, the average of the normalized projections to Ψ over $\Psi_{j\vec{k}}$. The maximum value of the average error is 1, as Ψ is normalized in norm 2 to that value; for orthogonal wavelets, $\in_j = 0$. Hence, $\in_j$ gives a measure of the error made by assuming that the wavelet is orthogonal. In Table 1 the values of $\in_j$ are given; we see that the errors are very small, really negligible, in most instances; however, there is still a significative overlap between scale 0 and scale 1 of about 10%. This deviation from perfect orthoghonality is probably causing the observed errors when reconstructing the series and the main point to be improved numerically.

j	1	2	3	4	5	6	7
$\in_j$	0.10	0.02	0.006	0.004	0.003	0.003	0.002

Table 1
Average error of the orthogonality condition for the optimal wavelet

4.2 Representation of images with the wavelet basis

In Figure 5 we show the expansion up to different resolution levels for two images (shown in Figure 4) expanded in the wavelet basis. One of the images belongs to our ensemble and the other does not, but the performance is similar in both cases.

We define the efficiency of the wavelet representation as the ratio of the norm 2 of the wavelet coefficients to the norm 2 of the signal, that is, it is the square root of the ratio of the energy of the representation to the energy of the original signal (see eq. (43) for the precise definition). The efficiencies ϵ_1 for those two images are 0.75 (*imk03236.imc*) and 0.59 (Lena), which are rather far from 1.0 (perfect retrieval). It is evident that a significant amount of the information is lost in this kind of representation. It becomes even more evident for *imk03236.imc* as it is dominated by a vertical structure (vertical

edges) which has almost no overlap with the wavelet (which is aligned with the horizontal axis).

Fig. 4. Sample images *imk03236.imc* and Lena image

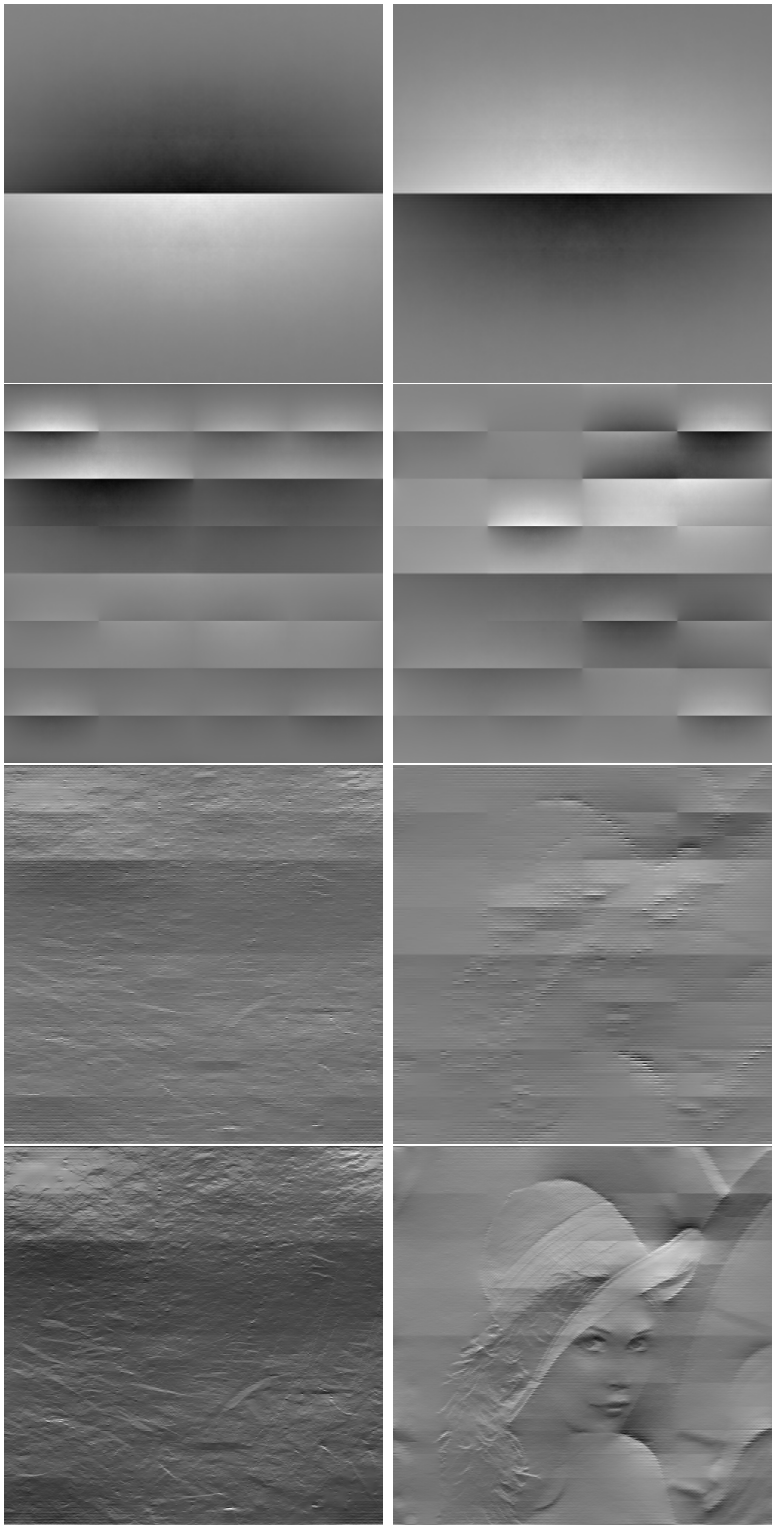

Fig. 5. $\sum_{j\vec{k}} \alpha_{j\vec{k}} \Psi_{j\vec{k}}(\vec{x})$ for $j = 0$, $j \leq 2$, $j \leq 6$ and $j \leq 8$ for *imk03236.imc* and Lena image.

5 The necessity for orientation: the example of Haar basis

So far, we have shown that the multifractal mother wavelet is not enough to provide a complete description of natural scenes. We claim that this lack of completeness is due to the orientational character of the basis wavelet, which makes necessary to introduce a sufficient number of rotated versions of the multifractal wavelet to achieve a perfect representation of images. In fact, the model we propose is quite simple: we will try to expand the images as the addition of the wavelet series associated to each oriented wavelet. In order to justify this approach, we present in this section analogous results for the well-known Haar basis (Daubechies, 1992).

The one-dimensional Haar wavelet $H(x)$ is given by the following, simple mathematical expression:

$$H(x) \;=\; \begin{cases} 1 & \text{if } \; 1/2 < x < 1 \\ -1 & \text{if } \; 0 < x < 1/2 \\ 0 & \text{otherwise} \end{cases} \tag{19}$$

The wavelet projections on the Haar wavelet splits the area of projection in two parts and returns the difference of the averages over each part. It is easy to verify that, under dyadic scaling and integer displacement the resulting wavelets are orthonormal, that is,

$$\langle H_{jk} | H_{j'k'} \rangle \;=\; 2^{-j}\, \delta_{jj'}\, \delta_{kk'} \tag{20}$$

It can also be proven that the dyadic Haar wavelet basis is a complete, orthonormal basis in the space of 1D functions of zero average (see (Daubechies, 1992) for the proof). So, the straightforward generalization of the Haar basis for 2D could consist in just adding one additional dimension to the 1D surrogate, keeping the value constant along that line. We define the horizontal Haar wavelet H^1 as:

$$H_1(x_1, x_2) \;=\; H(x_1)\chi_{[0,1]}(x_2) \tag{21}$$

where χ_A is a function valued 1 over the set A and zero outside. The horizontal Haar wavelet is represented in figure 6

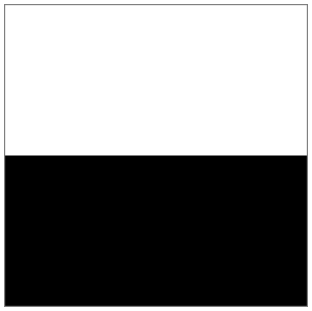

Fig. 6. Horizontal Haar mother wavelet H_1. Black: -1, white: +1

The horizontal Haar basis defines an orthonormal dyadic wavelet basis, but unfortunately this basis is not complete. As it is shown in figure 8, the horizontal Haar basis acts much as the multifractal wavelet, mainly describing horizontal features (horizontal edges), but being almost unconcerned by vertical features. In fact, for the case of Haar function, the solution to overcome the difficulty is well known: it is necessary to extend the concept of wavelet representation, including other "voices" or mother wavelets, capable to take account of different features. For the case of Haar basis, the other two voices are normally chosen as the vertical Haar wavelet H_2

$$H_2(x_1, x_2) \;=\; \chi_{[0,1]}(x_1)\, H(x_2) \tag{22}$$

and the diagonal Haar wavelet H_3:

$$H_3(x_1, x_2) \;=\; H(x_1)H(x_2) \tag{23}$$

as represented in Figure 7.

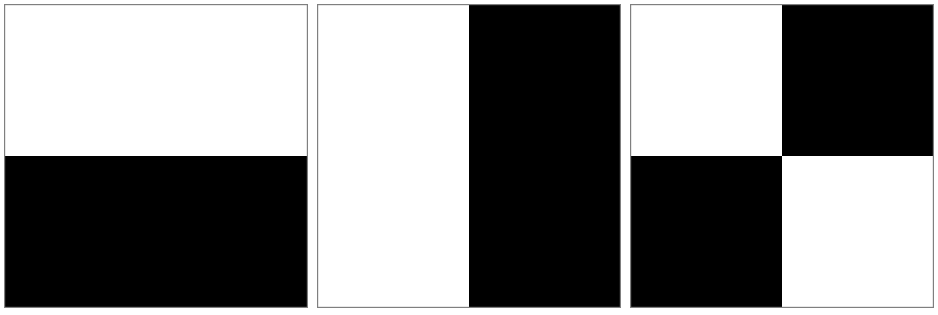

Fig. 7. Complete 2D Haar basis. Black: -1, white: +1

It can be proven (see again (Daubechies, 1992)) that this three-voice dyadic wavelet basis is an orthonormal, complete basis, that is:

$$\langle H_{rj\vec{k}} | H_{r'j'\vec{k}'} \rangle \;=\; 2^{-2j} \delta_{rr'}\, \delta_{jj'}\, \delta_{\vec{k}\vec{k}'} \tag{24}$$

and every zero-mean signal $c(\vec{x})$ can be represented exactly and uniquely as:

$$c(\vec{x}) \;=\; \sum_{rj\vec{k}} \alpha_{rj\vec{k}}\, H_{rj\vec{k}}(\vec{x}) \tag{25}$$

The examples for the representation on the three-voice dyadic basis are shown in Figure 8. We have also shown the representation in the two-voice dyadic basis given by the horizontal and vertical wavelets to stress the fact that for images dominated by the horizontal-vertical statistics this basis provides a very good approximation; however, in general the contributions of the diagonal wavelet are still necessary to correctly retrieve the image, specially if it is more isotropic.

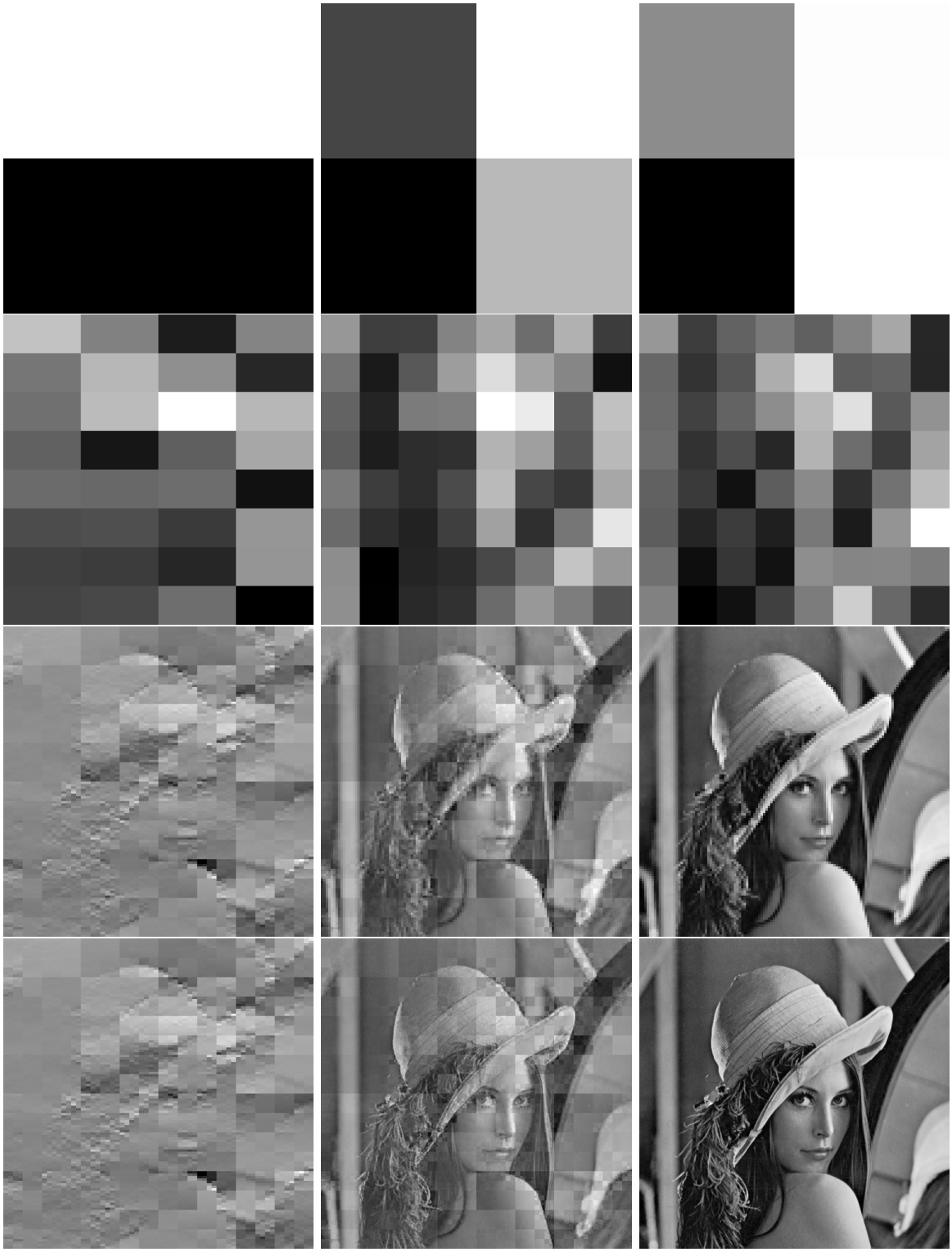

Fig. 8. $\sum_{r=1}^{n} \sum_{j\vec{k}} \alpha_{rj\vec{k}} H_{rj\vec{k}}(\vec{x})$ for $j = 0$, $j \leq 2$, $j \leq 6$ and $j \leq 8$ for Lena image and different number of basis functions n. From left to right: $n = 1$, 2 and 3

Let us remark that Haar basis is not a good candidate to provide an optimal code. The *wavelet coefficients* obtained with this basis have been observed to possess statistical dependencies (Huang and Mumford, 1999). Since these dependencies appear for any wavelet, they give no evidence about the independence of the coefficients η. In fact, the coefficients $\eta_{rj\vec{k}}$ associated to the Haar basis are not independent levels of resolution. We performed an experimental verification of this by computing the mutual information (Cover and Thomas, 1991) between $\eta_{rj\vec{k}}$ and $\eta_{r,j+1,2\vec{k}}$ over an subensemble of 100 random images on the same database used to learn the optimal wavelet. To increase sampling statistics we assumed traslational and scale invariances. We found that each Haar coefficient tree has a mutual information of about 0.01 bits between adjacent scales, while for the optimal basis the value was negligibly small (smaller than 10^{-4} bits (Turiel and Parga, 2000c)) [2] .

The example of Haar basis just evidences the necessity of introducing new voices in the wavelet scheme. It is very remarkable that Haar wavelets acts much like horizontal, vertical and diagonal edge detectors.We present now the generalization of the derivation of the formula determining the optimal wavelet to the case of a multi-voice, orientational wavelet expansion.

6 Oriented bases

The optimal wavelet basis introduced in (Turiel and Parga, 2000b) is very appealing from the point of view of image representation: it provides variables $\eta_{j\vec{k}}$ which are organized in independent levels of resolution. Such a representation could allow separating the relevant informative structures seen at each scale. However, as it has been shown in Section 4, this basis is far from being complete: the wavelet expansion, eq. (4), just describes horizontal features (horizontal edges). This is not so astonishing if we take into account that the experimental wavelet (Figure 2) has a clear, sharp transition aligned with an horizontal line. On the other hand, the discussion in the previous Section shows that it is possible to generalize eq. (4) to include a new degree of freedom; even more, in the case of the Haar basis, the new degree of freedom labels an orientation-like feature. Our guess is that the optimal wavelet basis derived so far is undercomplete; trying to extract just one feature detector, only the most frequent one in the learning set was obtained. The goal is to obtain a complete basis, using an expansion including a new degree of freedom

[2] To improve the assessment of independency of the optimal wavelet in contrast with Haar basis it would be necessary to study the variable $\log|\eta_{rj\vec{k}}|$, because $\eta_{rj\vec{k}}$ is very concentrated in small values and gives poor numerical accuracy; however much more statistics is necessary.

(which would somehow label the orientation of the feature detector) and at the same time satisfying the requirement of optimality issued in the previous derivation.

We will be forced to make some assumptions in order to solve this problem. In the rest of this section, we will show that the previously derived optimal wavelet is just a combination of the new feature detectors. We will also assume that those feature detectors are just rotated versions of one of them, and that they are mutually orthogonal; those conditions are not necessary to prove the connection with the former optimal wavelet, but they are essential to extract the feature detectors from the combination, as explained in Section 7. Finally, it must be experimentally checked that the new basis is orthogonal, complete and provides independent levels of resolution (independence among scales, but not necessarily among positions and/or orientations); we devote Section 8 to the presentation of the experimental results.

We now extend the previous theoretical analysis to the case in which there is no longer a single mother wavelet but a discrete collection of different wavelets, $\{\phi_r\}_{r=0}^{n-1}$. Each wavelet ϕ_r is tuned to capture features oriented according to a given preferred direction and to ignore the same features when they are missaligned with respect to the wavelet. We assume that each ϕ_r is a rotated version of $\phi_0 \equiv \phi$ and that they define an orthogonal basis. More precisely, let us define the rotation operator $\mathbf{R}$ which transforms the vector $\vec{x}$ into the vector $\mathbf{R}\vec{x}$, with the same modulus as $\vec{x}$ but making an angle of $\frac{2\pi}{n}$ radians with it. The operator $\mathbf{R}$ acts over ϕ in the way:

$$\mathbf{R}\phi(\vec{x}) \;\equiv\; \phi(\mathbf{R}\vec{x})$$

Hence, we will assume that:

$$\phi_r(\vec{x}) \;\equiv\; \mathbf{R}^r\phi(\vec{x}) \;=\; \phi(\mathbf{R}^r\vec{x}) \quad \text{and} \quad \langle\phi_r|\phi_{r'}\rangle \;=\; \langle\phi|\phi_{r'-r}\rangle = 0 \;\; \forall r \neq r' \tag{26}$$

Due to the cyclic character of the operator $\mathbf{R}$ ($\mathbf{R}^n = Id$), in what follows we will always regard any index r as an element of Z_n, which means that expressions such as $r + r'$ have to be understood "*modulus* n". The reader is specially warned to bear that in mind for the sums.

Analogously to eq. (4), we expand the contrast $c(\vec{x})$ as:

$$c(\vec{x}) \;=\; \sum_{r=0}^{n-1} \sum_{j=0}^{\infty} \sum_{\vec{k}\in(Z_{2^j})^2} \alpha_{rj\vec{k}}\, \phi_{rj\vec{k}}(\vec{x}) \tag{27}$$

Now we suppose that a relation analogous to eq. (8) holds for each $\eta_{rj\vec{k}}$ for fixed r.

We suppose that each tree of wavelet coefficients has in general nothing to do with the others, although independency among the $\eta_{rj\vec{k}}$'s for different r's is not required. Besides, we assume that the $\eta_{rj\vec{k}}$ follow the same distribution for all $r, j, \vec{k}$. This assumption is supported by the evidence presented in (Turiel and Parga, 2000a); in fact, they have the same distribution as $\eta_{j\vec{k}}$. Anyway, the condition could be relaxed to the requirement of equivalence of their first order moments.

In what follows, we will refer to the multifractal wavelet Ψ presented before (which was obtained under the assumption of isotropy) as the "isotropic wavelet" (even if in practice the resulting wavelet is not at all isotropic). The new wavelet ϕ expanding the rotational basis will be referred to as "oriented wavelet". It can be shown that the isotropic wavelet can be expressed as a linear combination of the oriented wavelets, namely:

$$\Psi \;=\; \sum_{r=0}^{n-1} p_r \phi_r \tag{28}$$

for some unknown weights p_r. The details of the proof can be found in Appendix A.

Some remarks are in order:

- Although it will be necessary for the following derivations, we have not made any assumption about the definition of the ϕ_r's nor about their possible orthogonality. So, this equation would also be the starting point if any of those hypotheses were changed.
- The distributions of the coefficients $\eta_{rj\vec{k}}$ are all the same, but the global coefficients $\alpha_{r0\vec{0}}$ which expand the different rotational pyramids are not. Had they all been the same, then Ψ could have been expressed as a sum of rotated wavelets, each with the same weight, so in particular $\mathbf{R}\Psi = \Psi$. But as Ψ does not exhibit such property, the weights p_r must be non-uniform. In fact, under more general assumptions they could even be non-positive. As we have the freedom of choosing the norm of ϕ, we fix it by choosing a convenient normalization for the weights, namely $\sum_r p_r^2 = 1$ (making use of orthogonality this means $\langle\Psi|\Psi\rangle = \langle\phi|\phi\rangle$).
- According to eq. (28) the isotropic wavelet Ψ can be expanded over the linear space defined by ϕ and its rotated versions. Due to the linearity and cyclic character of the operator R, we can obtain a $n \times n$ matrix relating the vector formed by Ψ and its rotated versions with the same type of vector expanded by ϕ. The inverse of that matrix, if it exists, will express ϕ as a linear combination of Ψ and its rotated versions.

7 Linear determination of the orthogonal oriented basis

In eq. (28) everything on the r.h.s. is undetermined: we know neither the oriented mother wavelet ϕ nor the weights $\{p_r\}$. We will face this problem by assuming that each rotation of ϕ is orthogonal to the original wavelet. On the one hand, this represents a physically more realistic case (misaligned features will not be detected). On the other hand, if there exists at least one such orthogonal solution, any other linear solution is a linear combination of that one. The existence or non existence of such a solution is determined by the invertibility of the matrix given by the scalar products $\langle \mathbf{R}^r \Psi | \mathbf{R}^{r'} \Psi \rangle$.

Let us pose the problem. We define the n-dimensional vectors $\vec{a} = (a_r)$ and $\vec{p} = (p_r)$, and the n-dimensional vector functions $\vec{\phi} = (\mathbf{R}^r \phi)$ and $\vec{\Psi} = (\mathbf{R}^r \Psi)$. Eq. (28) is then expressed as:

$$\Psi \;=\; \vec{p} \cdot \vec{\phi} \tag{29}$$

while we are looking for the vector $\vec{a}$ verifying the inverse relation:

$$\phi \;=\; \vec{a} \cdot \vec{\Psi} \tag{30}$$

The action of the operator $\mathbf{R}$ over vectors is defined by the relation:

$$\mathbf{R}\Psi \;=\; \mathbf{R}\vec{p} \cdot \vec{\phi} \tag{31}$$

which is obtained by just rewriting eq. (28). When expressed in coordinates, the previous expression corresponds to:

$$(\mathbf{R}\vec{p})_r \;=\; p_{r-1} \tag{32}$$

Using this expression and eqs. (29) and (30), the vectors $\vec{p}$ and $\vec{a}$ must be such that:

$$\sum_{r'} a_{r-r'}\, p_{r'} \;=\; \delta_{r0} \tag{33}$$

It is convenient to make use of discrete Fourier transforms to simplify the relations. The Fourier transform of the vector $\vec{a}$ is another vector $\hat{\vec{a}}$ which is defined by:

$$\hat{a}_r \;\equiv\; (\hat{\vec{a}})_r \;=\; \frac{1}{\sqrt{n}} \sum_{r'} a_{r'}\, e^{-2\pi i\, r' r/n} \tag{34}$$

and analogously for the inverse Fourier transform, changing the sign in the imaginary exponential. It can be checked that the latter is the true inverse of the discrete Fourier transform. So eq. (33) reads:

$$\hat{a}_r \hat{p}_r \;=\; 1 \tag{35}$$

that is, the elements of $\hat{\vec{p}}$ are the inverse of the elements of $\hat{\vec{a}}$. What is needed now is to determine the vector $\vec{a}$. For that, we will make use of the fact that $\{\phi_r\}$ is an orthogonal basis. Let us define the matrix $\mathbf{G}^\Psi$ of scalar products of the vector $\vec{\Psi}$, given by the matrix elements:

$$G^\Psi_{rr'} \;=\; \langle \mathbf{R}^r \Psi | \mathbf{R}^{r'} \Psi \rangle \tag{36}$$

Actually, $G^\Psi_{rr'} \;=\; \langle \Psi | \Psi \rangle \; g_{|r-r'|}$ where

$$g_r \;=\; \frac{\langle \Psi | \mathbf{R}^r \Psi \rangle}{\langle \Psi | \Psi \rangle} \tag{37}$$

and we will also write $\vec{g} \;=\; (g_r)$. Analogously, we can define the matrix $\mathbf{G}^\phi$ of scalar products of the vector $\vec{\phi}$; but due to the orthogonality $G^\phi_{rr'} = \langle \phi | \phi \rangle \; \delta_{rr'}$. Using eq. (30) and $\langle \phi | \phi \rangle = \langle \Psi | \Psi \rangle$ (due to our normalization), we obtain the following equation:

$$\delta_{rr'} \;=\; \sum_{nn'} a_{n-r} a_{n'-r'} g_{n-n'} \tag{38}$$

which is nothing but a standard decorrelation or orthogonality relation. By means of discrete Fourier transforms this relation reads:

$$1 \;=\; |\hat{a}_r|^2 \; \hat{g}_r \tag{39}$$

The vector $\vec{g}$ can be calculated from the isotropic wavelet Ψ, which we already know. Hence, the Fourier transform of the vector $\vec{a}$ can be computed by means of the formula:

$$\hat{a}_r \;=\; \hat{g}_r^{-1/2} \; e^{i\Phi_r} \tag{40}$$

which also defines ϕ using eq. (30). Finally, combining eqs. (40) and (35) we obtain the expresion for the Fourier transform of the weight vector $\vec{p}$,

$$\hat{p}_r \;=\; \hat{g}_r^{1/2} \; e^{-i\Phi_r} \tag{41}$$

Remarks:

- The existence of the inverse vector $\vec{a}$ depends on the nature of the coefficients $\hat{g}_r$. By construction, they are necessarily real numbers, but need not to be positive or non zero. There will exist an inverse if and only if $\hat{g}_r > 0 \ \forall r$.
- The existence of well-defined symmetry or antisymmetry axes in the wavelet should also be reflected in the structure of the weight vector $\vec{p}$.
- We have a large degree of freedom in the construction of the inverse vector $\vec{a}$, namely the choice of the phases $\{\Phi_r\}$, that is, the choice of a unitary transformation. We can limit somewhat this arbitrariness recalling that $\vec{a}$ must be a real vector; thus $\Phi_{n-r} = -\Phi_r$. Anyway we have still to choose $n/2$ phases.

To solve the undetermination on the phases $\{\Phi_r\}$, we have made the simplest choice for them, namely $\Phi_r = 0 \ \forall r$. This corresponds to the local solution, i.e., the one which is the most concentrated around $r = 0$. This means that the weight p_0 is maximum in this solution, and thus the corresponding oriented wavelets ϕ will be the most similar to the isotropic Ψ. What is somewhat surprising is that up to $n = 8$, $p_0 = 1$ and $p_r = 0$ for $r > 0$ with an error of less than 2%, what means that in practice $\Psi = \phi$. This is not surprising because the statistics of images is left-right symmetric but not up-down symmetric (the top of the images is usually clearer because the sky appears usually at that part of the scene); so, in the sum given in eq. (28) the vertical edge-detectors tend to cancel (a vertical edge is equally likely to be produced by a transition from a dark left side to a light right side or by the opposite transition), while the horizontal edge-detector will survive (because the bottom is darker in average than the top).

So Ψ itself can be used to expand the oriented basis. However, an additional information is still required: the number of different orientations n. A way to estimate this is to compare the energy of the image with that of its coefficients; if c can be expanded as in eq. (27) it follows that:

$$\langle c^2 \rangle \ = \ \sum_{r=0}^{n-1} \sum_{j=0}^{\infty} \sum_{\vec{k} \in (Z_{2^j})^2} 2^{-2j} (\alpha_{rj\vec{k}})^2 \tag{42}$$

This only holds if the wavelet basis $\{\phi_{rj\vec{k}}\}_{rj\vec{k}}$ is an orthogonal, complete basis (eq. (42) is just Parseval's relation). Provided that the basis is orthogonal but not complete, the equality does not longer hold, but it can be used to estimate n. We define the efficiency of the wavelet representation, $\epsilon_n[c]$, as the square root of the ratio of the energy of the coefficients to the energy of the image (in physical terms these so-called "energies" are in

fact powers of luminance flux),

$$\epsilon_n[c] \;=\; \sqrt{\frac{\sum_{rj\vec{k}} 2^{-2j}(\alpha_{rj\vec{k}})^2}{\langle c^2 \rangle}} \tag{43}$$

The closer $\epsilon_n[c]$ is to 1 for a given image c the better this image is represented by the wavelet expansion (the critical assumption here is the orthogonality of the wavelet basis). If $\epsilon_n[c]$ is greater than 1 the wavelets are oversampling the image and the representation is necessarily redundant (which in turn implies that orthogonality cannot longer hold). Assuming that each one of the different orientations affords a similar quantity of energy, the efficiency provides a quantitative measure of the redundancy in this case.

8 Experimental results on oriented bases

Figures 9 and 10 show the representation at $n = 2$ and 3 of the same two images. As a small modification with respect to the previously presented theory, the elementary angle is π/n instead of $2\pi/n$ because $\Psi(-\vec{x}) = -\Psi(\vec{x})$ is not independent of $\Psi(\vec{x})$. We are restricting our attention to the space of antisymmetric functions and this fact implies making several technical modifications with respect to the derivation presented in Section 7: the operator R is redefined multiplying the previous definition by a factor $e^{i\pi/n}$ to fulfill the condition $R^n = Id$ (to assure cyclicity on the index r) and the vectors $\vec{a}$ and $\vec{p}$ are now complex vectors, the normalization on $\vec{p}$ being expressed as $\sum_{r=0}^{n-1} |p_r|^2 = 1$. The derivation is essentially the same, taking care of introducing complex conjugates when appropriate (for instance, in the scalar products). Anyway, what is experimentally observed is that $\sum_{r=1}^{n-1} |p_r|^2 < 0.02$ up to $n = 8$, so with great precision $\phi = \Psi$.

As the wavelets were obtained assuming orthogonality, we should check this property. It is unnecessary to check the orthogonality between the different j's and $\vec{k}$'s for fixed r: as $\phi = \Psi$; it was already verified for ϕ_0 in Section 4 and due to the rotational invariance of the inner product, it verifies also for the other ϕ_r's. It is then only necessary to check orthogonality between different orientations. Analogously to eq. (18), we define the average error as:

$$\in_{nj} \equiv \sum_{\vec{k}} |\langle \phi | R_n \phi_{j\vec{k}} \rangle| \tag{44}$$

where R_n equals to the rotation operator of angle π/n. The values of the average errors for $n = 1$, $n = 2$ and $n = 3$ are given in Table 2. They should be exactly equal to 0 for

all r and j, except for $r = 1$ and $j = 0$ since $\in_{10} = |\langle\phi|\phi\rangle| = 1$. We observe that for $n = 2$ the wavelets are close to orthogonality; however, for $n = 3$ there is a small coupling for several scales j. We will assume that orthogonality holds for $n = 2$ and that it is just an approximation for $n = 3$.

j	0	1	2	3	4	5	6	7
$\in_{1j}$	1.000	0.106	0.023	0.006	0.004	0.003	0.003	0.002
$\in_{2j}$	0.001	0.013	0.005	0.001	0.001	0.000	0.000	0.000
$\in_{3j}$	0.085	0.028	0.037	0.025	0.013	0.007	0.003	0.002

Table 2
Average error of the orthogonality condition for the optimal wavelet at $n = 1$, 2 and 3.

As in (Turiel and Parga, 2000b) and (Turiel and Parga, 2000c), the hypothesis of independence among scales should be self-consistently validated. Let us notice that the hypothesis only requires independence between $\eta_{rj\vec{k}}$ and $\alpha_{r,j-1,[\frac{\vec{k}}{2}]}$ at every scale j, location $\vec{k}$ and orientation r. We have checked this independence measuring the mutual informations between $\eta_{rj\vec{k}}$ and $\alpha_{r,j-1,[\frac{\vec{k}}{2}]}$ for a subensemble of 100 images, assuming translational invariance to increase sampling. The calculated mutual informations were smaller than 10^{-3} bits at all scales j and two orientations ($n = 2$) for a maximum of 11 bits.

A direct consequence of the independence of $\eta_{rj\vec{k}}$ from $\alpha_{r,j-1,[\frac{\vec{k}}{2}]}$ is that eq. (8) is well-behaved (that is, dividing $\alpha_{rj\vec{k}}$ by $\alpha_{r,j-1,[\frac{\vec{k}}{2}]}$ will never give rise to an infinity)[3] . This property follows from the fact that, under the requirement of independence, the distribution of $\eta_{rj\vec{k}}$ can be identified with that of the multiplicative process in eq. (3). But multiplicative processes on finite variation signals are bounded (Turiel and Parga, 2000a), that is, $\eta_{rj\vec{k}}$ has a maximum, finite value. For that reason, when $\alpha_{r,j-1,[\frac{\vec{k}}{2}]}$ is very small, $\alpha_{rj\vec{k}}$ also and the ratio is kept finite. It could be argued that large amounts of additive noise could eventually make eq. (8) ill-behaved; however, the existence of additive noise implies a violation of the multiscaling property, which has been extensively verified (Turiel et al., 1997, 1998; Turiel and Parga, 2000a; Turiel et al., 2000; Nevado et al., 2000). . We have experimentally observed that tbe value of $\eta_{rj\vec{k}}$ is bounded, reinforcing its validity.

In Table 3 the values of the efficiencies for the two images are provided. The results for other images are similar. It is clear that one orientation ($n = 1$) is always insufficient to describe the images correctly. On the other hand, $n = 3$ representations always oversamples by a considerable amount, around 50% of the power of image, and sometimes greater (as

[3] This result is in contrast with the case for general, non-optimal wavelets, which can give rise to relatively large values of $\alpha_{rj\vec{k}}$ while $\alpha_{r,j-1,[\frac{\vec{k}}{2}]}$ are small, making the determination of $\eta_{rj\vec{k}}$ from eq. (8) ill-behaved.

for Lena image). The best choice seems to be $n = 2$. The efficiency is close to 1, but always a bit above it. The cause for this excess of power in the wavelet representation is probably the error committed by assuming orthogonality in the wavelet representation. We think that the error could be disminished by improving the accuracy of the determination of the mother wavelet (for instance, increasing the learning set).

	1	2	3
imk03236.imc	0.75	1.06	1.45
Lena	0.59	1.17	1.75

Table 3
Efficiencies $\epsilon_n[c]$ at different orientational representations for images *imk03236.imc* and Lena

Let us notice that those results have been validated for an ensemble of images with fixed quantization noise and spatial resolution. However, the multiplicative process has been extensively verified for very different ensembles at different quantization noises and resolutions (Turiel et al., 1998; Nevado et al., 2000; Turiel et al., 2000; Turiel and Parga, 2000b). The existence of a multiplicative process (essential for the determination of the optimal wavelet) is then robust with respect to those effects.

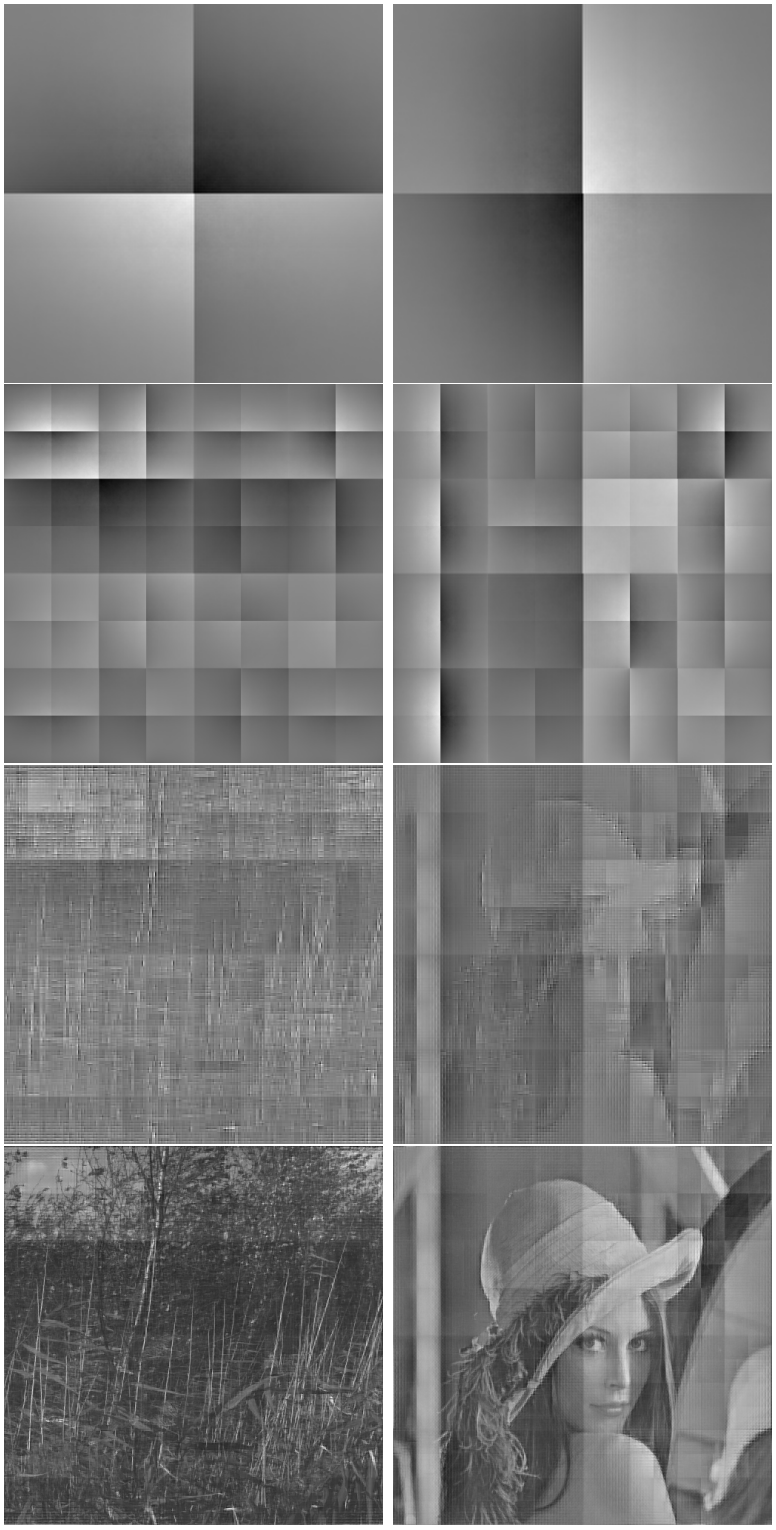

Fig. 9. $\sum_{rj\vec{k}} \alpha_{rj\vec{k}} \Psi_{rj\vec{k}}(\vec{x})$ for $j = 0$, $j \leq 2$, $j \leq 6$ and $j \leq 8$ for *imk03236.imc* and Lena image with $n = 2$ orientations

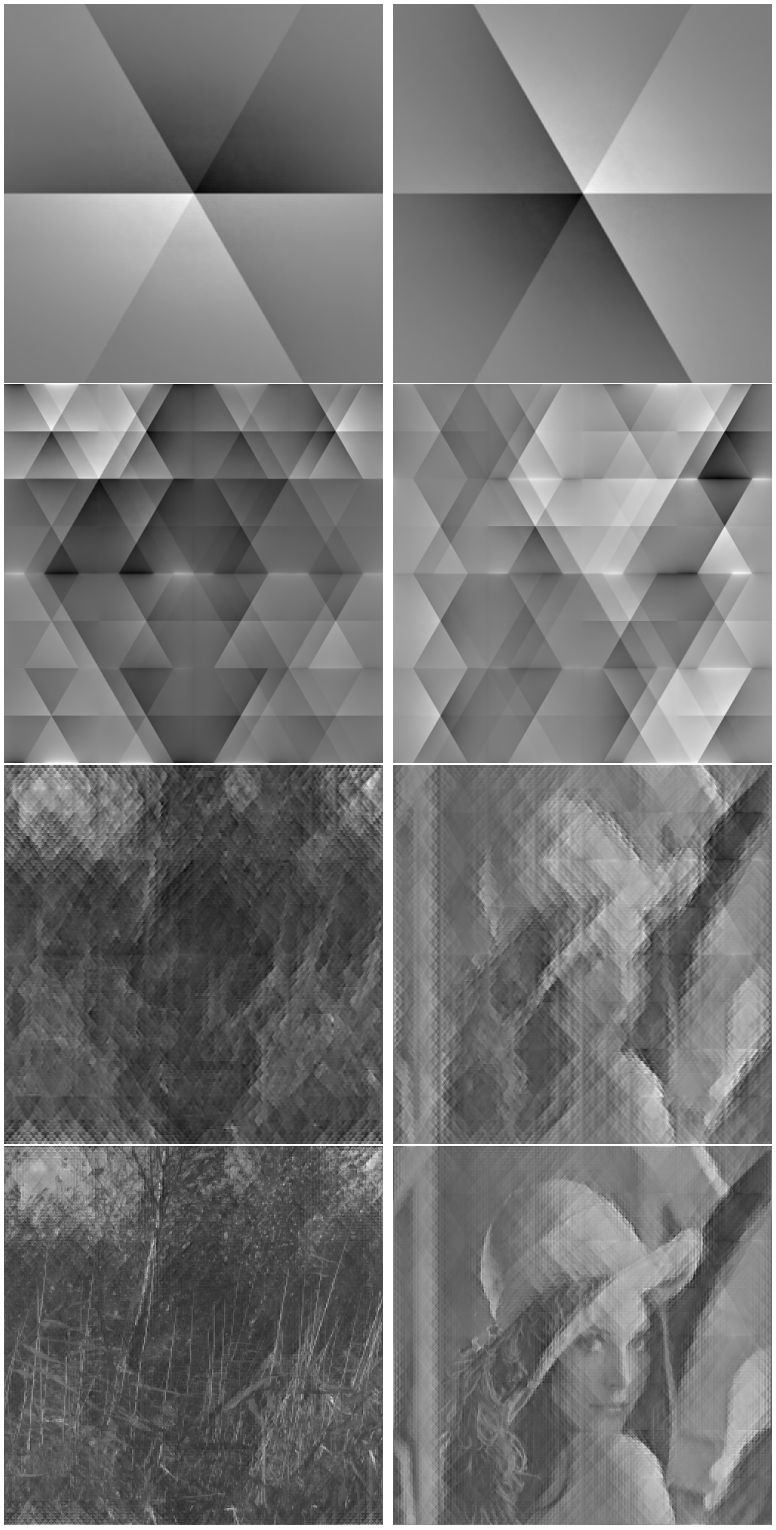

Fig. 10. $\sum_{rj\vec{k}} \alpha_{rj\vec{k}} \Psi_{rj\vec{k}}(\vec{x})$ for $j = 0$, $j \leq 2$, $j \leq 6$ and $j \leq 8$ for *imk03236.imc* and Lena image with $n = 3$ orientations

9 Discussion

In this paper, we have first reviewed the concept of optimal wavelet and we have discussed its capabilities for representing natural images. The optimal wavelet can be deduced as a consequence of the multiscaling properties which have been observed in natural images (reviewed in Sec. (2)). By construction, the optimal wavelet allows a multiscaling representation, in which the image is decomposed in independent levels of resolution. Independency had already been checked in previous work, but the completeness of such a representation had not been assessed. In fact it is rather straightforward to show that images are incompletely represented by this wavelet basis.

A study of the properties of the wavelet puts in evidence how the representation works and, more importantly, how it can be generalized to provide a complete representation. The optimal wavelet acts as an edge detector at each scale, but only along the horizontal direction. The example of the Haar basis makes it plausible that a complete representation could be obtained from a finite number of rotated versions of a non-isotropic wavelet. The derivation we have presented in this work shows that the previously, isotropically calculated wavelet can be used to expand the oriented basis in a small number of rotated versions with a good aproximation. It is important to notice that with this wavelet independency of each level of resolution is automatically granted.

The derived optimal wavelet has remarkable properties: edge detection in a finite number de rotated directions, transparency to the power spectrum, extraction of independent levels of resolution. All these properties are observed in the first levels of visual processing in humans and other mammals. The optimal wavelet is thus a good candidate for modelling visual information processing in the brain.

Research directions to be addressed in future works concern the dependency relations at a given level of resolution (spatial dependency, dependency among orientations). The understanding of those relations will allow to extract the independent components of this coding scheme, i.e., to provide a truly efficient code. A different and somewhat more important issue is to obtain the minimal number of orientations. One approach is to proceed along the same line as for the derivation of the isotropic wavelet, working under a constraint of independency in the representation. Even better, instead of restricting the wavelets to be rotated versions of the same function, it should be possible to construct a basis in a number of "voices" (different mother wavelets, necessary to provide a complete representation). This would allow, on the one hand, to assess if orientation can be deduced as a necessity for optimal codes; on the other hand, to deal with the finite resolution (sampling directions by neural cells). Once a complete, optimal code were accesible by means of these techniques, overcomplete, sparse representations (Olshausen and Field, 1996), (Olshausen and Field, 1997) could be explored, this would be more plausible for

a real biological coding - in particular some redundancy is required to insure stability of the representation against noise and small displacements of the image (Simoncelli et al., 1992).

Acknowledgements

A. Turiel is financially supported by a post-doctoral fellowship from INRIA. This work has been partially funded by the French-Spanish Picasso collaboration program (00-37) and by a French (DGA 96 2557A/DSP) and a Spanish grant (BMF2000-0011).

References

Atick, J. J., 1992. Could information theory provide an ecological theory of sensory processing? Network: Comput. Neural Syst. 3, 213–251.

Atick, J. J., Redlich, A., 1990. Towards a theory of early visual processing. Neural Computation 2, 308–320.

Atick, J. J., Redlich, A., 1992. What does the retina know about natural scenes? Neural Computation 4, 196–210.

Barlow, H. B., 1961. Possible principles underlying the transformation of sensory messages. In: Rosenblith, W. (Ed.), Sensory Communication. M.I.T. Press, Cambridge MA, p. 217.

B.Barlow, H., 1994. What is the computational goal of the neocortex? In: Koch, C., Davis, J. (Eds.), Large Scale Neuronal Theories of the Brain. MIT Press, Cambridge, MA, Ch. 1, pp. 1–22.

Bell, A. J., Sejnowski, T. J., 1997. The independent components of natural scenes are edge filters. Vision Research 37, 3327–3338.

Buccigrossi, R. W., Simoncelli, E. P., 1999. Image compression via joint statistical characterization in the wavelet domain. IEEE Transactions in Image Processing 8, 1688–1701.

Carandini, M., Heeger, D. J., Movshon, J. A., 1997. Linearity and normalization of simple cells of the macaque primary visual cortex. J. of Neuroscience 17, 8621–8644.

Cover, T. M., Thomas, J. A., 1991. Elements of information theory. John Wiley, New York.

Dan, Y., Atick, J. J., Reid, R. C., 1996. Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory. J Neurosci 16, 3351–3362.

Daubechies, I., 1992. Ten lectures on wavelets. CBMS-NSF Series in Ap. Math. Capital City Press, Montpelier, Vermont.

Feller, W., 1966. An introduction to probability theory and its applications. Volume 2. Wiley, New York.

Field, D. J., 1987. Relations between the statistics of natural images and the response properties of cortical cells. J. Opt. Soc. Am. 4, 2379–2394.

Field, D. J., 1994. What is the goal of sensory coding? Neural Computation 6, 559–601.

Heeger, D., 1992. Normalization of cell responses in cat striate cortex. Visual Neuroscience 9, 181–198.

Huang, J., Mumford, D., 1999. Statistics of natural images and models. In: Proc. CVPR. pp. 541–547.

Li, Z., Atick, J. J., 1994. Towards a theory of the striate cortex. Neural Computation 6, 127–146.

Linsker, R., 1988. Self-organization in a perceptual network. Computer 21, 105.

Maffei, L., Fiorentini, A., 1973. The visual cortex as a spatial frequency analyser. Vision Research 13, 1255–1267.

Mallat, S., Zhong, S., 1992. Characterization of signals from multiscale edges. IEEE Trans. on Pattern Analysis and Machine Intelligence 14, 710–732.

Nadal, J.-P., Parga, N., 1994. Nonlinear neurons in the low-noise limit: a factorial code maximizes information transfer. Network: Computation in Neural Systems 5, 565–581.

Nevado, A., Turiel, A., Parga, N., 2000. Scene dependence of the non-gaussian scaling properties of natural images. Network 11, 131–152.

Novikov, E. A., 1994. Infinitely divisible distributions in turbulence. Physical Reviev E 50, R3303.

Olshausen, B., Field, D. J., 1997. Sparse coding with an overcomplete basis set: A strategy employed by v1? Vision Research 37, 3311–3325.

Olshausen, B. A., Field, D. J., 1996. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381, 607–609.

Ruderman, D. L., 1994. The statistics of natural images. Network 5, 517–548.

Simoncelli, E. P., Freeman, W. T., Adelson, E. H., Heeger, D. J., 1992. Shiftable multiscale transforms [or "what's wrong with orthonormal wavelets"]. IEEE Trans Information Theory, Special Issue on Wavelets 38(2), 587–607.

Turiel, A., del Pozo, A., 2002. Reconstructing images from their most singular fractal manifold. IEEE Trans. on Im. Proc. 11, 345–350.

Turiel, A., Mato, G., Parga, N., Nadal, J. P., 1997. Self-similarity properties of natural images. In: Proc. of NIPS'97. Vol. 10. MIT Press, pp. 836–842.

Turiel, A., Mato, G., Parga, N., Nadal, J. P., 1998. The self-similarity properties of natural images resemble those of turbulent flows. Physical Review Letters 80, 1098–1101.

Turiel, A., Parga, N., 2000a. The multi-fractal structure of contrast changes in natural images: from sharp edges to textures. Neural Computation 12, 763–793.

Turiel, A., Parga, N., 2000b. Multifractal wavelet filter of natural images. Physical Review Letters 85, 3325–3328.

Turiel, A., Parga, N., 2000c. Wavelet based decomposition of natural images in indepen-

dent resolution levels. In: Pajunen, P., Karhunen, J. (Eds.), Proceedings of the Second International Workshop on Independent Component Analysis and Blind Signal Separation (ICA2000). Helsinki, Finland, pp. 339–344.
Turiel, A., Parga, N., Ruderman, D., Cronin, T., 2000. Multiscaling and information content of natural color images. Physical Review E 62, 1138–1148.
van Hateren, J. H., 1992. Theoretical predictions of spatiotemporal receptive fields of fly lmcs, and experimental validation. J. Comp. Physiology A 171, 157–170.
van Hateren, J. H., van der Schaaf, A., 1998. Independent component filters of natural images compared with simple cells in primary visual cortex. Proc. R. Soc. Lond. B265, 359–366.
Wainwright, M. J., Simoncelli, E. P., 2000. Scale mixtures of gaussians and the statistics of natural images. In: Solla, S. A., Leen, T. K., Müller, K.-R. (Eds.), Adv. Neural Information Processing Systems, vol. 12. MIT Press, pp. 855–861.
Wainwright, M. J., Simoncelli, E. P., Willsky, A. S., 2001. Random cascades on wavelet trees and their use in modeling and analyzing natural images. Applied Computational and Harmonic Analysis 11, 89–123.
Wornell, G. W., 1993. Wavelet-based representations for the 1/f family of fractal processes. Proc. of the IEEE 81, 1428–1450.

A Relation between the isotropic and the oriented wavelet

We repeat somewhat the original derivation for the isotropic wavelet (Turiel and Parga, 2000b). Using eq. (27), we obtain the following expression for the Fourier transform of the normalized contrast [4] :

$$\hat{C}(\vec{f}) \;=\; \sum_{r=0}^{n-1} \overline{|\alpha_{r0\vec{0}}|} \sum_{j\vec{k}} \left(\overline{|\eta_r|}\right)^j \; e^{-2\pi i\, 2^{-j}\vec{k}\cdot\vec{f}}\, 2^{-2j} \hat{\phi}_r(2^{-j}\vec{f}) \tag{A.1}$$

Defining the dilation operator $T_a F(\vec{f}) \equiv F(a\vec{f})$ we obtain:

$$\hat{C}'(\vec{f}) \;=\; \sum_{r=0}^{n-1} \overline{|\alpha_{r0\vec{0}}|} \sum_{j} \left(\frac{\overline{|\eta_r|}}{4}\right)^j \; T_{2^{-j}} \hat{\phi}'_r(\vec{f}) \tag{A.2}$$

[4] Recall that the normalization is chosen such that the signs of the wavelet coefficientes do not contribute any longer.

where $\hat{\phi}'_r \equiv \frac{\hat{\phi}_r}{\Lambda}$, $\hat{C}' \equiv \frac{\hat{C}}{\Lambda}$. Defining the new variable $\vec{v} = \log_2 \vec{u} \equiv (\log_2 u_1, \log_2 u_2)$ and $\hat{I}(\vec{v}) \equiv \hat{C}'(2^{\vec{v}})$, $\hat{F}_r(\vec{v}) \equiv \hat{\phi}'_r(2^{\vec{v}})$, it follows:

$$\hat{I}(\vec{v}) \;=\; \sum_{r=0}^{n-1} \overline{|\alpha_{r0\vec{0}}|} \sum_j \left(\frac{\overline{|\eta_r|}}{4}\right)^j \hat{F}_r(\vec{v} - j(1,1)) \tag{A.3}$$

Introducing the functions $\hat{\Omega}_r(\vec{v}) \equiv \overline{|\alpha_{r0\vec{0}}|} \sum_j \left(\frac{\overline{|\eta_r|}}{4}\right)^j \delta(\vec{v} - j(1,1))$, the previous equation is reduced further to:

$$\hat{I}(\vec{v}) \;=\; \sum_{r=0}^{n-1} \hat{\Omega}_r \otimes \hat{F}_r(\vec{v}) \tag{A.4}$$

where $\otimes$ stands for the convolution product. If we assume now that $\overline{|\eta_r|} = \overline{|\eta|}\ \forall r$, defining $p_r = \overline{|\alpha_{r0\vec{0}}|}/\overline{|\alpha_{0\vec{0}}|}$ the last equation can be rewritten as:

$$\sum_{r=0}^{n-1} p_r \hat{F}_r(\vec{v}) \;=\; \frac{1}{\overline{|\alpha_{0\vec{0}}|}} \left[\hat{I}(\vec{v}) \;-\; \frac{\overline{|\eta|}}{4}\, \hat{I}(\vec{v} - (1,1))\right] \tag{A.5}$$

Changing the variables $\vec{v}$ to variables $\vec{u}$, and comparing the r.h.s. with eq. (9), we arrive to the final relation between the rotational and the isotropic bases, eq. 28