Learning efficient internal representations from natural image collections

Antonio Turiel†, José M. Delgado‡ and Néstor Parga‡*

† Departament de Física Fonamental. Universitat de Barcelona.
Diagonal, 647. 08028 Barcelona. Spain
‡ Departamento de Física Teórica. Universidad Autónoma de Madrid
28049 Cantoblanco, Madrid. Spain

Abstract

Learning in sensory systems takes place after a repeated exposure to the incoming signals. Many theories based in information theoretical principles have been proposed to explain the synaptic adaptation which improves the coding capabilities of sensory areas. In this paper we show that a simple, natural learning rule leads to the known multifractal filter in natural images, which is used to split images into independent components by resolution levels in a non-linear way. The result shows the biological plausibility of this coding strategy not only in the visual pathway but also in other sensory modalities.

Keywords: learning, coding, statistical analysis, wavelets.

1 Introduction

There has been much work recently at the interphase between the modelling of visual systems and computer vision. This is justified because on the one hand a better understanding of natural systems may lead to new image processing algorithms, and on the other hand the analysis natural image properties may help to model the early visual system of mammals.

Based on early works of Barlow [1], many works have focussed on the use of information theoretic concepts in order to address the question of efficiency of neural coding. A possible efficiency criterium is to maximize the information transfer [2]. As shown in [3], 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 these ideas of information transfer and redundancy reduction [4], [5], [6], [7]. Any representation should arrive to a compromise between scale and translationnal invariances (that is, a multiscale representation) as they cannot be exactly fulfilled at the same time [8]. Direct statistical analysis of natural images leads also naturally to a multiscale analysis, see [9, 10, 11].

In previous studies, a multifractal analysis of natural images has been performed on a wide variety of ensembles of natural images [12]. It has been shown that an optimal wavelet [13] can be constructed (learned) from a set of images. It is however necessary to introduce oriented wavelets (with exactly two orientations) to provide a complete representation [14]. The representation so obtained achieves both whitening and edge detection. More importantly, the representation based 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. In this paper we will show that a simple learning rule, which corresponds to a Hebb-like synaptic matrix with lateral inhibitions, gives rise to that optimal wavelet representation.

*To whom correspondence should be addressed

2 Learning rule

Consider a learning set consisting of N natural images, each one represented by their local intensity contrast $c_i(\vec{x})$ with $i = 1, \ldots N$ and the vector $\vec{x}$ giving the position of the point on the image. A simple learning rule which would give a maximum response to the images in the learning set would be a Hebb-like synaptic matrix. Hence, the filter F^N applied to any image $I(\vec{x})$ (not necessarily in the learning set) would be given by:

$$F^N[I] \equiv \int d\vec{x}\; m^N(\vec{x}) I(\vec{x}) \tag{1}$$

where $m^N(\vec{x})$ is the mean of the local contrast of the images in the learning set,

$$m^N(\vec{x}) \equiv \frac{1}{K^N} \sum_{i=1}^{N} c_i(\vec{x}) \tag{2}$$

where the normalization constant K^N is computed as: $K^N = (\sum_{i,j} \int d\vec{x}\; c_i(\vec{x}) c_j(\vec{x}))^{1/2}$. For perfectly orthogonal learning images, the filter F^N will give maximum reponses over the learning set. On the contrary, if the learning images consist of orthogonal subparts (independent objects), the filter F^N will tend to give maximum response to those objects.

The advantages of a filter like F^N is that it can be learnt in an accumulative way, following Hebb's rule; this property makes it also biogically plausible, as the synaptic strength is updated according to the sequence of activities (proportional to the values of luminosity) in the sensory network. The filter F^N is defined over images of a given size, and it provides a scalar, continuous output for images of that size. In order to produce a code for images, one should apply the filter at different sizes and different locations in the image, in order to detect the independent objects present at different scales and locations. This would imply to learn one filter for each scale j and location $\vec{k}$. However, this is not really necessary: it is known [15] that the statistical ensemble of natural images is translationally and scale invariant. This statement means that the filters F^N learnt at different locations are the same, provided that the learning sets are large enough (N sufficiently large); it also means that the filter F^N learnt from images of a given size is a resized version of the filter F^N learnt from images of different size. This allows simplifying the computational task, as we just need to compute F^N at a single size and location.

The backdraw with the approach above is that the filter F^N is not optimal: it could be responding to features already detected at smaller scales in different locations contained in the receptive field of F^N. Hence, once a filter $F^N_{j\vec{k}}$ at scale j and location $\vec{k}$ is responding to the presence of a features, then all the $F^N_{j'\vec{k}'}x$ are also responding provided that j' corresponds to a scale greater than j and the filter at position $\vec{k}'$ overlaps the filter at position $\vec{k}$. This kind of behaviour is associated to the so-called "edge persistency" in wavelet transforms [16].

There is a way to overcome this difficulty, and it is to define new filters F^N_Ψ given by the substraction from F^N of the contributions of the other F^N at smaller scales. Analogously to eq. (1) the filter F^N_Ψ operates over a given image $I(\vec{x})$ as

$$F^N_\Psi[I] \equiv \int d\vec{x}\; \Psi^N(\vec{x}) I(\vec{x}) \tag{3}$$

and the function Ψ^N is defined by differences of means m^N at different scales. One of the simplest filters Ψ^N is the so-called "dyadic multifractal wavelet" [13], which is given by

$$\Psi^N(\vec{x}) = \frac{1}{\mathcal{N}} \left[m^N(\vec{x}) - \frac{1}{2} \sum_{l_1, l_2 = 0,1} m^N(2\vec{x} - \vec{l}) \right] \tag{4}$$

where $\mathcal{N}$ is a normalization constant. Looking at eq. (4), we can see that the density for the filter F^N is compared to the four contributions of filters F^N defined at linear scales exactly $\frac{1}{2}$ smaller; this is the reason of the factor $\frac{1}{2}$

in the equation (it is related to some scaling properties and the translational invariance, see [11, 13]). Eq. (4) is schematically represented in Figure 1, for a single image ($N = 1$).

Let us emphasize that the filter F_Ψ^N can also be learned accumulatively, as the expression which defines it is linear in the learning images c_i. In a more biological context, it implies a correction to a simple Hebb's rule, in which neurons associated to smaller scales act inhibitorily over neurons in the following scale, inhibiting their reponse to features which were already detected at smaller sizes and forcing them to concentrate in the truely new arriving features.

$c(\vec{x})$ $\quad\quad$ $\sum_{\vec{l}\in Z^2} c(2\,\vec{x}\,-\,\vec{l})$ $\quad\quad$ Ψ^1

Figure 1: Construction of the filter F_Ψ from a single image

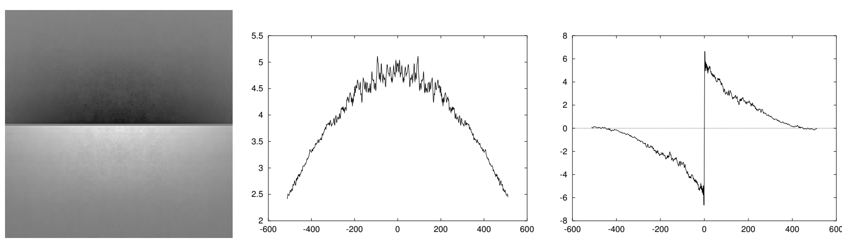

Figure 2: Left: Gray level representation of the filter learnt from 4000 images (white: positive values, black: negative values); Middle: Horizontal cut; Right: vertical cut

3 Optimal filter as a wavelet basis

The filter F_Ψ is equivalent to a wavelet projection [17] on the wavelet Ψ. If Ψ defines a orthogonal wavelet basis, any image I can be reconstructed from those projections, as a linear combination of the wavelet projections with the wavelet Ψ resized at each scale j and relocated at each position $\vec{k}$, in the way:

$$I(\vec{x}) \;=\; \sum_j \sum_{\vec{k}} \; \alpha_{j\vec{k}}\; \Psi_{j\vec{k}}(\vec{x}) \tag{5}$$

where $\alpha_{j\vec{k}} \;=\; \int d\vec{x}\;\; \Psi_{j\vec{k}}(\vec{x}) I(\vec{x})$. But in general, Ψ is not a single feature, but a combination of a number n of them, in the way:

$$\Psi \ = \ \sum_{r=0}^{n} \phi_r \tag{6}$$

and the general wavelet expansion is:

$$I(\vec{x}) \ = \ \sum_{r=0}^{n-1} \sum_{j=0}^{\infty} \sum_{\vec{k} \in (Z_{2^j})^2} \alpha^r_{j\vec{k}} \, \phi^r_{j\vec{k}}(\vec{x}) \tag{7}$$

Notice that eq. (7) does not imply that the basis $\{\phi^r_{j\vec{k}}\}$ is strictly complete: it may happen to be overcomplete, introducing a small degree of redundancy, useful to control noise.

4 Properties of the filters

The wavelets defined by eqs. (4) and (6) have been studied in ([13, 18, 14]). This type of wavelets possess very remarkable properties, the most important being that they define optimal filters from the perspective of efficient coding. This property can be expressed as a multiplicative relation:

$$\alpha^r_{j\vec{k}} \ = \ \eta^r_{j\vec{k}} \, \alpha^r_{j-1\left[\frac{\vec{k}}{2}\right]} \tag{8}$$

where the variables $\eta^r_{j\vec{k}}$ are independent from the $\alpha^r_{j-1\left[\frac{\vec{k}}{2}\right]}$ and have the same distribution for all the wavelet indices r, resolution levels j and spatial locations $\vec{k}$. The variables $\eta^r_{j\vec{k}}$ contain the "novelties" encountered when passing from a large scale level (indexed $j-1$) to a smaller one (indexed by j). Thus, a system coding only the variables "of scale change" $\eta^r_{j\vec{k}}$ would be retaining just independent variables, supressing unnecessary redundancies and optimizing the coding cost. Let us remark that obtaining the variables of scale change implies a *non-linear processing* of the wavelet coefficients $\alpha^r_{j\vec{k}}$. Optimal linear coding of natural images (e.g. as in [6]) is not possible, in particular because it is not able to deal with the redundancy coming from the persistency of edges across scales.

Furthermore, due to the multiplicative character of eq. (8) the distribution of $\eta^r_{j\vec{k}}$ is necessarily highly kurtotic ([10, 19]). Kurtotic distributions are interesting from the perspective of coding [7], as they imply a smaller cost in coding (the majority of the coefficients have negligible values, and need not to be coded). Kurtosis is also another manifestation of sparseness, a property of neural coding [20]. Notice however that kurtosis is this model is not introduced as an external requirement (as in [7]), but *it is a consequence of the statistical properties of images*, with no prior.

A technique to obtain each feature wavelet ϕ_r from the combination Ψ has been proposed in [14]. It was assumed that every ϕ_r is just a rotated version of the same wavelet ϕ. The experimentally obtained ϕ defines an almost ortoghonal pyramid, that is, the autoprojections $\langle\phi|\phi_{j\vec{k}}\rangle$ are negligible for $j \neq 0$, $\forall\vec{k}$. The values of those autoprojections are very small, about 1%, except those for $j=1$, which are about 10%. We think that the projections would become smaller using larger training ensembles, and this is still a point to be improved (also, more general feature filters (not simply rotated versions) could improve the performance). For $N=2$ rotations, the two pyramids are mutually orthogonal, $\langle\phi^r_{j\vec{k}}|\phi^{r'}_{j'\vec{k}'}\rangle \approx 0$ if $r \neq r'$ (error less than 1% in any instance, [14]). Hence, the basis with two oriented wavelets (horizontal and vertical) acts as an orthonomal basis. The coefficients $\alpha^r_{j\vec{k}}$ of any image c are then easily extracted by simple projection, $\alpha^r_{j\vec{k}} \ = \ \langle\phi^r_{j\vec{k}}|c\rangle$

Figure 3 shows the reconstruction at different levels of resolution of two example images. The coefficients $\alpha^r_{j\vec{k}}$ were extracted (assuming orthonormality) by just projecting on $\phi^r_{j\vec{k}}$. This introduces a significant error (the PSNRs for the reconstructions are 22.02 dB and 25.43 dB), although the main features are well described. From the figure it can be observed that the image is regenerated by the succesive addition of horizontal and vertical small lines. So, the wavelet representation codes the image as edges and contours, and the wavelet coefficients (which spawn the independent resolution levels) measure the relative illumination of such edges.

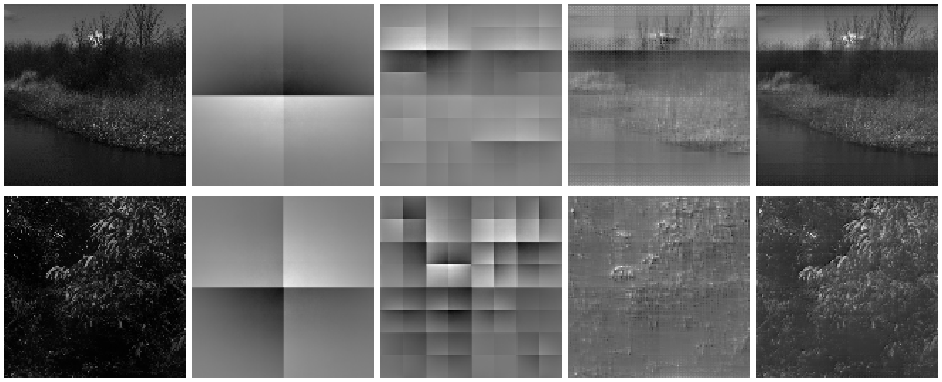

Figure 3: *From left to right:* Original image and $\sum_{rj\vec{k}} \alpha^r_{j\vec{k}} \Psi^r_{j\vec{k}}(\vec{x})$ for $j = 0$, $j \leq 2$, $j \leq 6$ and $j \leq 8$ with $n = 2$ orientations, for imk00480.imc (top) and imk02000.imc (bottom)

5 Discussion

In this work we have shown how to construct visual filters starting from rather general biologically plausible rules. The obtained filters mimic a Hebb-like learning rule which seems appropriate to describe images, but are still very redundant for coding. When some fundamental statistical properties of natural images are introduced, the learning rules are modified obtaining a simple, linear way to take account of redundancies between consecutive levels of resolution. The resulting filter turns out to be the known multifractal wavelet filter, which is known to be optimal for efficient coding.

Our presentation shows that simple principles, derived from *observational properties* of the statistics of signals, can give rise to coding schemes which are very efficient in both coding and processing. At the same time, the algorithms proposed seem to be strongly connected with the way in which biological systems act. We think that this methodology is not exclusive of visual system, and could be used to understand other sensory modalities.

Acknowledgements

This work has been partially funded by a Spanish grant (BMF2000-0011).

References

[1] H. B. Barlow, “Possible principles underlying the transformation of sensory messages,” in *Sensory Communication*, W. Rosenblith, Ed., p. 217. M.I.T. Press, Cambridge MA, 1961.

[2] R. Linsker, “Self-organization in a perceptual network,” *Computer*, vol. 21, pp. 105, 1988.

[3] J.-P. Nadal and N. Parga, “Nonlinear neurons in the low-noise limit: a factorial code maximizes information transfer,” *Network: Computation in Neural Systems*, vol. 5, pp. 565–581, 1994.

[4] J. H. van Hateren, "Theoretical predictions of spatiotemporal receptive fields of fly lmcs, and experimental validation," *J. Comp. Physiology A*, vol. 171, pp. 157–170, 1992.

[5] J. J. Atick, "Could information theory provide an ecological theory of sensory processing?," *Network: Comput. Neural Syst.*, vol. 3, pp. 213–251, 1992.

[6] A. J. Bell and T. J. Sejnowski, "The independent components of natural scenes are edge filters," *Vision Research*, vol. 37, pp. 3327–3338, 1997.

[7] B.A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by v1?," *Vision Research*, vol. 37, pp. 3311–3325, 1997.

[8] X.-G. Li, P. Somogyi, A. Ylinen, and G. Buzsáki, "The hippocampal CA3 network: an in vivo intracellular labelling study," *J. Comp. Neurol.*, vol. 339, pp. 181–208, 1994.

[9] D L Ruderman, "Designing receptive fields for highest fidelity," *Network*, vol. 5, pp. 147–156, 1994.

[10] A. Turiel, G. Mato, N. Parga, and J. P. Nadal, "Self-similarity properties of natural images," in *Proc. of NIPS'97*. 1997, vol. 10, pp. 836–842, MIT Press.

[11] A. Turiel and N. Parga, "The multi-fractal structure of contrast changes in natural images: from sharp edges to textures," *Neural Computation*, vol. 12, pp. 763–793, 2000.

[12] A. Nevado, A. Turiel, and N. Parga, "Scene dependence of the non-gaussian scaling properties of natural images," *Network*, vol. 11, pp. 131–152, 2000.

[13] A. Turiel and N. Parga, "Multifractal wavelet filter of natural images," *Physical Review Letters*, vol. 85, pp. 3325–3328, 2000.

[14] A. Turiel, J.-P. Nadal, and N. Parga, "Orientational minimal redundancy wavelets: from edge detection to perception," To appear in Vision Research, 2003.

[15] D. L. Ruderman and W. Bialek, "Statistics of natural images: Scaling in the woods," *Physical Review Letters*, vol. 73, pp. 814, 1994.

[16] Stéphane Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 2nd Edition, 1999.

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

[18] A. Turiel, J.-P. Nadal, and N. Parga, "An algorithm for image representation as independent levels of resolution," 2002, vol. 2415, pp. 1213–1218.

[19] A. Turiel, G. Mato, N. Parga, and J. P. Nadal, "The self-similarity properties of natural images resemble those of turbulent flows," *Physical Review Letters*, vol. 80, pp. 1098–1101, 1998.

[20] W. E. Vinje and J. L. Gallant, "Sparse coding and decorrelation in primary visual cortex during natural vision," *Science*, vol. 287, pp. 1273–1276, 2000.