An algorithm for image representation as independent levels of resolution

Antonio Turiel[1], Jean-Pierre Nadal[2], and Néstor Parga[3]*

[1] Air Project - INRIA. Domaine de Voluceau BP105
78153 Le Chesnay CEDEX, France
[2] Laboratoire de Physique Statistique, Ecole Normale Supérieure
24 rue Lhomond. 75231 Paris CEDEX 05. France
[3] Departamento de Física Teórica. Universidad Autónoma de Madrid
28049 Cantoblanco, Madrid. Spain

Abstract. Recently it has been shown that natural images possess a special type of scale invariant statistics (multiscaling). In this paper, we will show how the multiscaling properties of images can be used to derive a redundancy-reducing oriented wavelet basis. This kind of representation can be learnt from the data and is optimally adapted for image coding; besides, it shows some features found in the visual pathway.

Keywords: wavelets, statistical analysis, coding

1 Introduction

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

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 maximizing the information transfer. As shown in [2], 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 [3], [4]. 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 [5]. Direct statistical analysis of natural images leads also naturally to a multiscale analysis, see [6], [7].

In previous studies, a multifractal analysis of natural images has been performed on a wide variety of ensembles of natural images [8]. It has been shown that an optimal wavelet [9] can be constructed (learned) from a set of images. It is however necessary

---

* To whom correspondence should be addressed

to introduce oriented wavelets (with exactly two orientations) to provide a complete representation [10]. The representation so obtained 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.

2 Dyadic wavelet bases

We will represent a particular image by its intensity $I(\boldsymbol{x})$ (i.e., graylevel in digitized images) at every point $\boldsymbol{x}$ in the screen. For normalization convenience, we will work over the contrast $c(\boldsymbol{x}) \equiv I(\boldsymbol{x}) - I_0$ where the normalization constant I_0 is chosen such that the average of c over the screen vanishes.

The starting point of our approach is the so called dyadic wavelet expansion. In this type of expansion, the signal is represented in successive levels of detail, from the coarsest (large scales) to the finest (small scales), which are obtained by resizing and translating a family of functions (wavelets) $\{\phi^r\}_r$. It is called "dyadic" because from one level of resolution to the following, the scale is divided by a factor two. The largest scale is fixed as the unity, $1 = 2^0$, and then the j-th scale is 2^{-j}. Assuming that the scale is of the same order as the dispersion of the wavelet, it is possible to distinguish up to 2^j different blocks along each dimension (2^{2j} blocks in our case, as the images are bi-dimensional). Then, a dyadic wavelet expansion for $c(\boldsymbol{x})$ corresponds to the following mathematical expression:

$$c(\boldsymbol{x}) = \sum_{r=0}^{N-1} \sum_{j=0}^{\infty} \sum_{\boldsymbol{k} \in (Z_{2^j})^2} \alpha_{j\boldsymbol{k}}^r \, \phi_{j\boldsymbol{k}}^r(\boldsymbol{x}) \tag{1}$$

where $\phi_{j\boldsymbol{k}}^r(\boldsymbol{x}) \equiv 2^j \phi^r(2^j \boldsymbol{x} - \boldsymbol{k})$. For any particular family of wavelets $\{\phi^r\}_r$ for which the dyadic representation eq. (1) holds, the coefficients $\alpha_{j\boldsymbol{k}}^r$ must satisfy the following statistical relation, known as "multiplicative cascade relation" [7]:

$$\alpha_{j\boldsymbol{k}}^r \doteq \eta_{j\boldsymbol{k}}^r \, \alpha_{j-1\left[\frac{\boldsymbol{k}}{2}\right]}^r \tag{2}$$

where "$\doteq$" means that both sides have the same distribution, but they are not necesarily equal for any image and location $j, \boldsymbol{k}$. The variables $\eta_{j\boldsymbol{k}}^r$ are independent from the $\alpha_{j-1\left[\frac{\boldsymbol{k}}{2}\right]}^r$ and have the same distribution for all the wavelet indices r, resolution levels j and spatial locations $\boldsymbol{k}$. This statistical property is a consequence of the fact that natural images are multiscaling signals [7].

3 The optimal wavelet family

It is possible to determine an optimal family of wavelets under some hypothesis. The optimal family is determined by requiring the statistical equality in eq. (2) to hold *point-by-point*, that is, the equality is true for any image, resolution and location. This is a very strong statement that will allow us to compute Ψ, which is a linear combination

of the elements in the family. Under additional constraints on the rotational character of the family, the whole family can be calculated. It also allow us to extract the coefficients $\eta_{j\boldsymbol{k}}^r$ from the wavelet coefficients $\alpha_{j\boldsymbol{k}}$

$$\eta_{j\boldsymbol{k}}^r = \alpha_{j\boldsymbol{k}}^r / \alpha_{j-1[\frac{\boldsymbol{k}}{2}]}^r \tag{3}$$

These variables $\eta_{j\boldsymbol{k}}^r$ provide a representation in which each level of resolution is independent of the others: $\eta_{j\boldsymbol{k}}^r$ and $\eta_{j'\boldsymbol{k}'}^r$ are statistically independent for $j \neq j'$ [9]. It is possible then to propose a neural architecture able to extract the independent components of images (see figure 1).
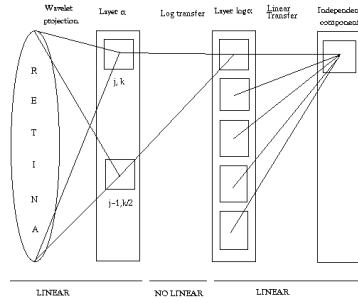

Fig. 1. A possible neural architecture to extract independent features using wavelet projection. The image is projected over $\phi_{j\boldsymbol{k}}^r$ to produce layer α, that is, the activity of each unit is proportional to the corresponding wavelet coefficient $\alpha_{j\boldsymbol{k}}^r$. Then, a logarithmic transformation is applied to produce layer $\log \alpha$. Finally, those activities are linearly combined to produce the independent components $\log \eta_{j\boldsymbol{k}}^r = \log \alpha_{j\boldsymbol{k}}^r - \log \alpha_{j-1,[\frac{\boldsymbol{k}}{2}]}^r$

The optimal wavelet family can be obtained from a set of images by means of an average of the contrast over all of them, $C(\boldsymbol{x})$. That function $C(\boldsymbol{x})$ does not give direct access to the family, but to an average wavelet Ψ, according to the following relation [9]:

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

where $\mathcal{N}$ is a normalization constant which is obtained by requiring that $\int d\boldsymbol{x} \, \Psi^2(\boldsymbol{x}) = 1$. In the case of having a one-wavelet family, Ψ equals the mother wavelet ϕ^0. In the more general case, as it is shown in [10], the general expression is:

$$\Psi(\boldsymbol{x}) = \sum_{r=0}^{N-1} p_r \, \phi^r(\boldsymbol{x}) \tag{5}$$

for some unknown weights p_r. Assuming that the different wavelets ϕ^r are rotated versions of the same function $\phi \equiv \phi^0$, it can be proven (see [10] for details) that a possible solution of eq. (5) is $p_r \approx \delta_{r0}$, that is, $\phi \approx \Psi$. This solution holds for different values of N, the number of different orientations. So, eq. (4) can be used to obtain the non-rotated wavelet $\phi \approx \Psi$ and to extract the whole family from it. In practice we will only consider the case $N = 2$, which was shown to be very close to an orthonormal basis.

4 Properties of the optimal wavelet

By construction, the wavelet family splits images in independent levels of resolution (the variables η_{jk}^r). It can be learnt on-line by simple addition of new images in the averages $C(\boldsymbol{x})$ appearing in eq. (4), which in turn is equivalent to average the learnt wavelets. The experimental wavelet ϕ is shown in Figure 2; it was learnt over an ensemble of 1000 images from H. van Hateren's web database (see [11] for details on their callibration).
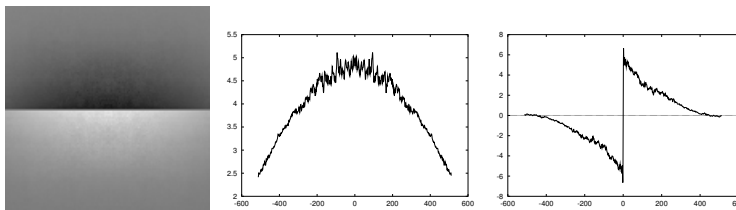

Fig. 2. Left: Gray level representation of the optimal wavelet ϕ (white: positive values, black: negative values); Middle: Horizontal cut; Right: vertical cut

The experimentally obtained ϕ defines an almost orthoghonal pyramid, that is, the autoprojections $\langle \phi | \phi_{jk} \rangle$ are negligible for $j \neq 0$, $\forall \boldsymbol{k}$. 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. For $N = 2$ rotations, the two pyramids are mutually orthogonal, $\langle \phi_{jk}^r | \phi_{j'k'}^{r'} \rangle \approx 0$ if $r \neq r'$ (error less than 1% in any instance, [10]). Hence, the basis with two oriented wavelets (horizontal and vertical) acts as an orthonomal basis. The coefficients α_{jk}^r of any image c are then easily extracted by simple projection, $\alpha_{jk}^r = \langle \phi_{jk}^r | c \rangle$

5 Representation

Figure 3 shows the reconstruction at different levels of resolution of three example images. The coefficients α_{jk}^r were extracted (assuming orthonormality) just projecting on ϕ_{jk}^r. This introduces a significant error (the PSNRs for the reconstructions are 22.02 dB, 25.43 dB and 22.63 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.
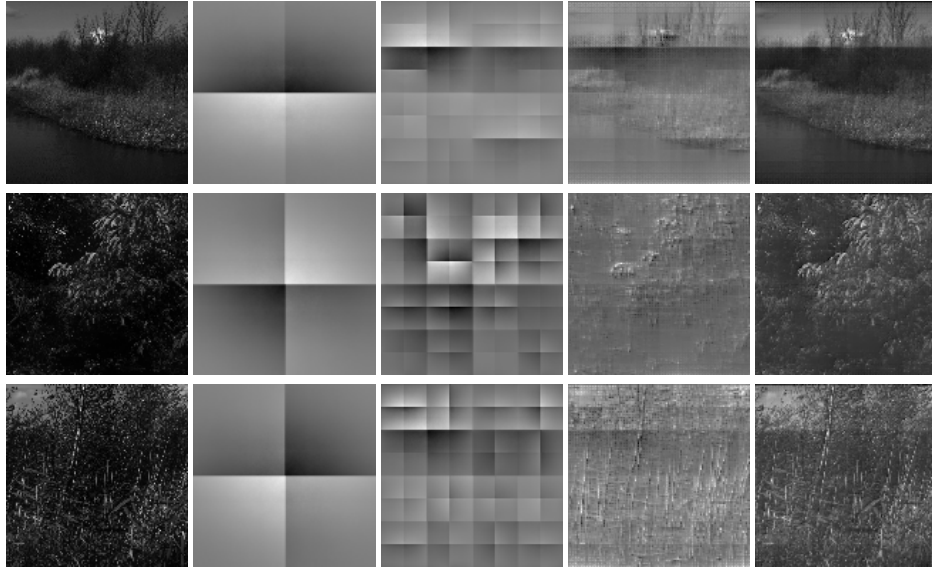

Fig. 3. *From left to right:* Original image and $\sum_{rjk} \alpha_{jk}^r \Psi_{jk}^r(x)$ for $j = 0$, $j \leq 2$, $j \leq 6$ and $j \leq 8$ with $n = 2$ orientations, for imk00480.imc (top), imk02000.imc (middle) and imk03236.imc (bottom)

6 Discussion

In this work we have seen that natural images possess an important property: they can be canonically represented in an optimal dyadic wavelet basis at two orientations. We have seen that such a basis can be expanded from a wavelet that can be calculated given a large enough learning set of images. The advantage of this particular representation lies in the fact that it naturally splits images in independent levels of resolution.

The proposed description of images as entities formed by independent levels of resolution has two main implications. First, it gives a justification of several features in the human visual system (learning capability, power spectrum extraction, edge detection), which are also present in this wavelet expansion. Our wavelet expansion could then be used to emulate a biological coding network. Second, the independency of the resolution levels is very useful to represent images compactly (independency reduces

coding cost with respect to other wavelet based codings, as the one presented in [12]). We think that the approach proposed here can be useful both to understand early areas in the visual cortex and to design good coding algorithms.

Acknowledgements

References

1. Barlow, H.B.: Possible principles underlying the transformation of sensory messages. In Rosenblith, W., ed.: Sensory Communication. M.I.T. Press, Cambridge MA (1961) 217
2. Nadal, J.P., Parga, N.: Nonlinear neurons in the low-noise limit: a factorial code maximizes information transfer. Network: Computation in Neural Systems **5** (1994) 565–581
3. van Hateren, J.H.: Theoretical predictions of spatiotemporal receptive fields of fly lmcs, and experimental validation. J. Comp. Physiology A **171** (1992) 157–170
4. Atick, J.J.: Could information theory provide an ecological theory of sensory processing? Network: Comput. Neural Syst. **3** (1992) 213–251
5. Li, X.G., Somogyi, P., Ylinen, A., Buzsáki, G.: The hippocampal CA3 network: an in vivo intracellular labelling study. J. Comp. Neurol. **339** (1994) 181–208
6. Ruderman, D.L.: The statistics of natural images. Network **5** (1994) 517–548
7. Turiel, A., Parga, N.: The multi-fractal structure of contrast changes in natural images: from sharp edges to softer textures. Neural Computation **12** (2000) 763–793
8. Nevado, A., Turiel, A., Parga, N.: Scene dependence of the non-gaussian scaling properties of natural images. Network **11** (2000) 131–152
9. Turiel, A., Parga, N.: Multifractal wavelet filter of natural images. Physical Review Letters **85** (2000) 3325–3328
10. A. Turiel, J.P.N., Parga, N.: Orientational minimal redundancy wavelets: from edge detection to perception. Submitted to Vision Research (2002)
11. van Hateren, J.H., van der Schaaf, A.: Independent component filters of natural images compared with simple cells in primary visual cortex. Proc. R. Soc. Lond. **B265** (1998) 359–366
12. Buccigrossi, R.W., Simoncelli, E.P.: Image compression via joint statistical characterization in the wavelet domain. IEEE Transactions in Image Processing **8** (1999) 1688–1701