Available online at www.sciencedirect.com
SCIENCE DIRECT®
NEUROCOMPUTING
Neurocomputing 69 (2006) 1224–1227
www.elsevier.com/locate/neucom

Receptive fields of simple cells from a taxonomic study of natural images and suppression of scale redundancy

José M. Delgado*, Antonio Turiel[1], Néstor Parga

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

Available online 2 February 2006

Abstract

Much effort has been carried out to propose models of the early visual pathway based on the statistical analysis of natural images. These conventional frameworks lead to predictions on the RFs of simple cells which do not fit well their observed properties [D.L. Ringach, Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex, J. Neurophysiol. 2002 (2002) 455–463]. In order to overcome these difficulties, we have been carrying a research program to derive robust coding principles from the statistics of natural images [A. Turiel, G. Mato, N. Parga, J.P. Nadal, Self-similarity properties of natural images, in: Proceedings of NIPS'97, vol. 10, MIT Press, 1997, pp. 836–842; A. Turiel, J.M. Delgado, N. Parga, Learning efficient internal representations from natural image collections, Neurocomputing 58–60 (2004) 915–921]. Two principles emerging from our study of image statistics are: first, there exists scale redundancy and this can be eliminated from the code; second, images are constructed by a combination (partly linear and partly non-linear) of some simple patterns of contrast (edges, bars, and composites of these two). These two principles can be used to derive filters which explain observed properties of the RFs of simple cells and which compare quite well with the results reported by Ringach [Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex, J. Neurophysiol. 2002 (2002) 455–463] and previous experimental work.

Keywords: Sensory coding; Learning; Feature detectors; Natural images

1. Introduction

It has been proposed that adaptation of a living organism to its habitat affects sensory coding in two complementary ways: by reducing the redundancy in the code [2] and by maximizing the information conveyed by the code about the stimuli [6], although both strategies are related [8,7]. These coding strategies have in common the necessity of adopting a statistical treatment of the sensory stimuli. Many studies have showed, first, that natural images have a non-trivial statistical structure and that much information can be obtained from their high-order statistics [13,14,18,19]; second, that knowledge about the statistics of images leads to precise predictions of the properties of cells in the visual pathway [1]. Many statistical models of the early visual pathway, and in particular of the receptive fields (RFs), have been proposed [1,3,4,10,15]. All these models are based on a priori assumptions about the statistical properties of the stimuli. However, their predictions on the RFs of simple cells do not fit well the observed properties [12].

Our approach consists of deriving coding principles from statistical properties of natural images, which could tell us as much as possible about the properties of the visual pathway and more particularly of the RFs of cells in primary visual cortex [11,17–21]. Our study conducted us to two main principles: first, there exists scale redundancy which can only be removed in a non-linear fashion [18,19]; second, images are constructed by a combination (partly linear and partly non-linear) of some simple patterns of contrast (edges, bars, and composites of these two) which are commonly found in images. We have used the filters

*Corresponding author. Tel.: +34 91 497 4884; fax: +34 91 497 3936.
E-mail addresses: josemaria.delgado@uam.es (J.M. Delgado), turiel@icm.csic.es (A. Turiel), nestor.parga@uam.es (N. Parga).
[1]Present address: ICM-CSIC, Passeig Maritim de la Barceloneta, 37-49. 08003 Barcelona, Spain

0925-2312/$ - see front matter
doi:10.1016/j.neucom.2005.12.081

derived with our model [17,21] to explain observed properties of the RFs of simple cells [11]. In this paper, we will present an explicit coding scheme based on these principle. In addition, we show some features of our predicted RFs which closely agree with the experimental evidence.

2. Scale redundancy and its implications in coding

Scale redundancy is an important property of natural images: it means that when the same scene is observed at two different scales (i.e., resolutions), the new details arising at the finer resolution appear in such a way that the total information content is still dominated by the coarser resolution features and hence both representations are very redundant. This property, which could seem evident, is rather subtle, as one can readily see by trying to construct a signal behaving in that way. In particular, scale redundancy implies that once a contrast modulation has been detected at a given scale of a natural image (what means that a contrast structure with neat, sharp transitions perceptible at the scale of detection has been segregated from its surround) it will be still detected at smaller scales. Although scale redundancy has been observed in many studies on the statistics of natural images [5,9,14,18,19], only recently this property has been explicitly used to derive filters which explicitly remove it from the code [21]. Our previous studies have shown that it is possible to obtain a complete basis of optimal filters, directly learned from ensembles of natural images and able to suppress scale redundancy [17,20], once some appropriate, global conditioning on the statistics is assumed (for instance, orthogonality among the basis elements). Our results showed that there is no need to impose properties such as edge-detection, orientation selectivity or sparse coding, as these emerge naturally from the predicted filters.

Learning an optimal filter $F_{\bar{I}}$ from a collection of images $\{I_n\}_{n=1,\ldots,N}$ requires the image prototype $\bar{I}$ of the selected image set, defined as

$$\bar{I} = \frac{1}{N}\sum_{n=1}^{N} I_n. \tag{1}$$

The filter is then constructed simply as a difference of the image prototype at two consecutive scales [21]:

$$F_{\bar{I}}(\vec{x}) = \bar{I}(\vec{x}) - \lambda \sum_{\vec{l}} \bar{I}(2\vec{x} - \vec{l}), \tag{2}$$

where the inhibitory terms $\bar{I}(2\vec{x} - \vec{l})$ are obtained by a scale reduction (here by a factor 2) applied to the prototype of the image set, which appear recentered so that the reduced images cover the same area as the original one (this is done with the displacement vectors $\vec{l}$). This formula differs from the one derived in [17,20,21] in that the strength of the inhibitory terms, λ, has been left free. It can then be adapted to the conditioning of the particular image collection being studied. The filters $F_{\bar{I}}$ suppress the scale redundancy, at least in the linear stage of the processing; but its output still needs to be processed non-linearly [18,19] to reach the actual infomax/redundancy suppression code [7,8].

To obtain a faithful representation of visual stimuli one should use a bank of filters which should be sufficiently large to produce a complete or overcomplete representation of the stimuli. Some possible banks can be derived using strong assumptions on the geometry of the filters [20]. A more reasonable, statistically consistent approach is to derive the filters from different image prototypes, each one learned from a different collection of images. We will call these image prototypes "image features", as they capture some relevant properties of small ensembles of images in which they are expressed as the salient contrast structure. Once these image features are known, the associated optimal filters have to be obtained with Eq. (2). Each filter is optimal for the image collection used to derive it, as was proved in [21]. This process can be repeated for collections of images at different scales, giving rise to multiscale filter banks; in practice, however, scaled versions of the basic prototypes appear at all scales [16].

3. Extraction of the relevant image features

In order to unveil the image features that explain the observed structure of natural images, it is necessary to perform an exhaustive taxonomic classification on a large dataset of natural images. Such a classification is very demanding in data and in computation time. We have done the feature classification using a large, well-known web database of images (see details on the images in [22]), but the complexity and depth of this work exceeds the limits of this article [16]. For the needs of the present work, let us just say that we have used a simple competition rule to split the ensemble in consistent image classes, each class dominated by a particular image feature. We have observed that the emerging features can be described in very simple terms: they are edges, bars or simple composite of these two basic types. In Fig. 1 we show an example of such a classification, performed over one image. Here, the ensemble consists of small blocks of 16×16 pixels (total: 6144 blocks) in which the whole image is decomposed, and the classification is performed over this ensemble. For this example, we show in Fig. 1 the five most represented classes (middle) and the associated image features (bottom).

To illustrate the appropriateness of using banks of features instead of individual features to derive optimal filters from Eq. (2), we have performed a simple test by approximating the image I by an image constructed from a linear combination of the image features. For a given prototype $\bar{I}_\alpha$, we derived the corresponding optimal filter $F_{\bar{I}_\alpha}$ ($\lambda = \frac{1}{2}$, as in [20]), and then we used it to represent the image as a combination of the response at each possible scale (in fact, a wavelet representation of the signal). For each feature $\bar{I}_\alpha$ we can thus define the approximation A_α, and we can also construct the optimal approximation A_{opt}