Chapter 12
Dependencies of Energy Detectors: Beyond V1

All the models in this book so far have dealt with the primary visual cortex (V1). In this chapter, we show how statistical models of natural images can be extended to deal with properties in the extrastriate cortex, i.e. those areas which are close to V1 (also called the striate cortex) and to which the visual information is transmitted from V1.

12.1 Predictive Modeling of Extrastriate Cortex

Most of the experimental results in early cortical visual processing have considered V1. The function of most extrastriate areas is still rather much a mystery. Likewise, most research in modeling natural image statistics has been on low-level features, presumably corresponding to V1.

However, the methodology that we used in this book could possibly be extended to such extrastriate areas as V2, V3(A), V4, and V5. Actually, since the function of most extrastriate areas is not well understood, it would be most useful if we could use this modeling endeavor in a predictive manner, so that we would be able to predict properties of cells in the visual cortex, in cases where the properties have not yet been demonstrated experimentally. This would give testable, quantitative hypotheses that might lead to great advances in visual neuroscience.

In the next sections, we attempt to accomplish such predictive modeling in order to predict properties of a third processing step, following the simple and complex cell layers. The predictions should be based on the statistical properties of modeled complex-cell outputs. Our method is to apply ordinary independent component analysis to modeled outputs of complex cells whose input consists of natural images.¹

12.2 Simulation of V1 by a Fixed Two-Layer Model

The basic idea in this chapter is to fix a model of complex cells and then learn a representation for complex cell outputs using a statistical model. The resulting three-layer network is depicted in Fig. 12.1.

This approach is rather different from the one used in previous chapters, in which we learned first the simple cells and then the complex cells from the data. Here, to simplify the model and the computations, we do not attempt to learn everything

¹This chapter is based on the article (Hyvärinen et al. 2005a), originally published in BMC Neuroscience. The experiments were done by Michael Gutmann. Copyright retained by the authors.
Fig. 12.1 The simplified hierarchical model investigated in this chapter. Modeled complex-cell responses are calculated in a feedforward manner, and these responses are subsequently analyzed by a higher-order feature layer in the network (“contour” layer). To emphasize that the lower layers are fixed and not learned, these layers have been greyed out in the figure. The direction of the arrows is from higher features to lower ones which is in line with the interpretation of our analysis as a generative model.

at the same time. Instead, we fix the first two layers (simple and complex cells) according to well-known models, and learn only the third layer.

The classic complex-cell model is based on Gabor functions. As explained in Sect. 3.4.2, complex cells can be modeled as the sum of squares of two Gabor functions which are in quadrature phase. Quadrature phase means simply that if one of them is even-symmetric, the other one is odd-symmetric. This is related to computation of the Fourier energy locally, as explained in Sect. 2.4.

Complex-cell responses $c_k$ to natural images were thus modeled with a Gabor energy model of the following form:

$$c_k = \left( \sum_{x,y} W_{o}^e(x, y)I(x, y) \right)^2 + \left( \sum_{x,y} W_{o}^o(x, y)I(x, y) \right)^2$$ (12.1)

where $W_{o}^e$ and $W_{o}^o$ are even- and odd-symmetric Gabor receptive fields; the equation shows that their squares (energies) are pooled together in the complex cell. The complex cells were arranged on a $6 \times 6$ spatial grid. They had $6 \times 6 = 36$ different spatial locations, and at each location, four different preferred orientations and three different frequency selectivities (“bands”). The aspect ratio (ratio of spatial length to width) was fixed to 1.5. The frequency selectivities of the Gabor filters are shown in Fig. 12.2, in which all the filters $W$ were normalized to unit norm for visualization purposes. The actual normalization we used in the experiments consisted of standardizing the variances of the complex cell outputs so that they were equal to unity for natural image input. The number of complex cells totaled $36 \times 4 \times 3 = 432$.

As the basic data, we used 1008 grey-scale natural images of size $1024 \times 1536$ pixels from van Hateren’s database. We manually chose natural images in the nar-