
Swiss Federal Institute of Technology Zurich

Diploma Thesis

Independent Components of Complex-Cell Outputs

Michael Gutmann

Spring 2004

Supervisor: Prof. Rodney Douglas
Institute of Neuroinformatics,
UNI-ETH Zurich

Adviser: Dr. Aapo Hyvärinen
Basic Research Unit,
Helsinki Institute for Information Technology

Abstract

An important approach in visual neuroscience is to investigate the relation between properties of the visual cortex and properties of its stimuli. It has been shown how classical receptive fields of simple and complex-cells in the primary visual cortex emerge from the statistical properties of natural images by forcing the cell responses to be maximally sparse and independent. A related, but more restrictive computational principle, non-negative sparse coding, has recently been shown to lead to the emergence of cells which code for contours of a given spatial frequency band. These cells were related to areas beyond the primary visual cortex.

In this thesis, we investigate how to learn cells beyond the primary visual cortex from the statistical properties of natural images by maximization of sparseness and independence only. We show that the non-negativity constraint is not a necessary condition. Further, we explain how maximization of independence and sparseness only can also lead to the emergence of cells which pool spatially coherent across-frequency activity from the primary visual cortex.

We propose that this kind of pooling allows the cells, in contrast to cells of the primary visual cortex, to code for realistic low level image features like for example composite edges.

Acknowledgments

I wish to thank Dr. Aapo Hyvärinen. For his willingness to become my adviser, for his way of giving advice and foremost, for the quality of his advice. It has been a pleasure to prepare this thesis under these circumstances. In an equal way, I like to thank Dr. Patrik Hoyer who has de facto become a co-adviser in a later stage of this thesis.

I appreciate that Prof. Rodney Douglas, Director of the Institute of Neuroinformatics, UNI-ETH Zurich, was willing to supervise this thesis. The thesis itself was prepared at the Basic Research Unit of the Helsinki Institute for Information Technology, a joint research institute of the University of Helsinki and the Helsinki University of Technology. I thank Research Director Prof. Heikki Mannila for providing the research environment.

Contents

Abstract	I
Acknowledgments	III
Table of Contents	V
1 Introduction	1
2 Background	5
2.1 Information theory	5
2.1.1 Entropy	5
2.1.2 Mutual information	6
2.1.3 Redundancy	6
2.2 Independent component analysis	8
2.2.1 Purpose	8
2.2.2 Information theoretic approach	8
2.2.3 Fixed-point algorithm	9
2.3 Natural images	10
2.3.1 Structure in natural images	10
2.3.2 Scale invariance	10
2.3.3 Non-Gaussianity	11
2.4 Primary visual cortex	11
2.4.1 Visual pathway	12
2.4.2 Models of cortical cells	12
2.4.3 Function of cortical cells	16
2.5 Edge detection	17
2.5.1 Difficulties in edge detection	17
2.5.2 Linear filtering approach	17

2.5.3	Quadratic filtering approach	18
3	Previous Work and Research Questions	21
3.1	Learning properties of the visual cortex	21
3.1.1	Theoretical framework	21
3.1.2	Analysis by synthesis	22
3.2	Learning simple-cell receptive fields	22
3.3	Learning complex-cell receptive fields	23
3.4	Learning cells beyond V1	25
3.5	Research questions	25
3.5.1	Standard ICA model of complex-cell outputs	25
3.5.2	The case of multiple scales	26
3.5.3	Validity of the experimental design	27
4	Standard ICA Model of Complex-Cell Outputs	29
4.1	Introduction	29
4.2	Methods	29
4.2.1	Experimental design	30
4.2.2	Free parameters	31
4.2.3	Interpretation of the ICA model	33
4.3	Results	35
4.3.1	Varying the learning criterion	35
4.3.2	Varying preprocessing	36
4.4	Conclusion	39
5	The Case of Multiple Scales	41
5.1	Introduction	41
5.2	Methods	41
5.3	Results: Existence of dependencies	42
5.4	Results: Higher-order cell properties	43
5.4.1	Generative weight structure - overview	45
5.4.2	Generative weight structure - by example	45
5.4.3	Higher-order cells in the network	48
5.5	Results: Significance of multiple scales	48
5.6	Conclusion	52

6 Validity of the Experimental Design	53
6.1 Introduction	53
6.2 Methods	53
6.2.1 Investigation by correlation analysis	53
6.2.2 Investigation by ICA	55
6.3 Results: Investigation by correlation analysis	56
6.4 Results: Investigation by ICA	57
6.5 Conclusion	62
7 Conclusions	63
7.1 Conclusions	63
7.2 Discussion	64
Bibliography	65

Chapter 1

Introduction

This thesis is about images and the way we perceive them. We will often mention natural images. These are images like the one shown in Figure 1.1. This image, for example, shows trunks of long trees which stand in a small area of grass surrounded by other trees. The trunks are fairly straight and they stand clearly out of the background forest. Not all natural images have exactly this structure, but they all depict wildlife scenes, e.g. landscapes as they have already existed thousands of years ago. That is, an image as in Figure 1.2 does not belong to this category.

However, what makes the former image different from the latter?

Clearly, we can perceive in the former well-defined objects such as trees, edges and lines, but in the latter, we have difficulties to detect any such structure. One reason for this is that the individual picture elements (pixels) in Figure 1.2 take their intensity values rather independently from each other, whereas in the natural image, the individual pixels work together to



Figure 1.1: Natural image.

Figure 1.2: Not a natural image.

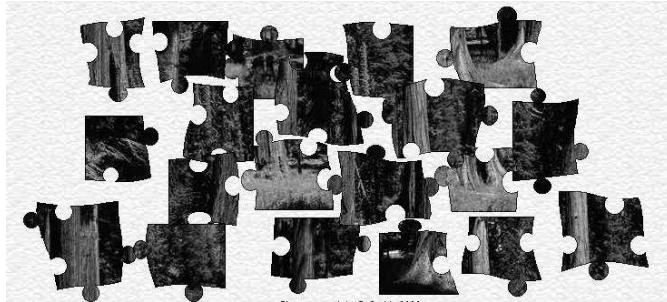


Figure 1.3: Exploring the structure in natural images.

give structure to the image. The individual pixel values depend on each other.

We do not need complicated analysis to find that Figure 1.2 lacks any structure as it can be found in Figure 1.1. We just *perceive* that pixel values in the natural image depend on each other. It happens without any effort thanks to our visual system. This system has apparently analyzed dependencies between the individual pixel values and detected structure in the image shown in Figure 1.1, not however in the image shown in Figure 1.2. This thesis is about how we explore structure in natural images.

To see what it takes to explore structure in images, let us solve the jigsaw puzzle in Figure 1.3. From our own experience, we might know that mechanical comparison of all the different picture elements with each other is not a very efficient strategy to discover the meaning of the image as given by the different picture elements. One might better first group picture elements locally together which show some similarities, then, leverage on that representation to yield iteratively groups of higher-order picture elements till the meaning of the image is discovered. That is, a possible strategy to explore the natural image could be to build it inside-out as illustrated in Figure 1.4.

It is likely that our visual system has found by evolution or development its own means to unveil the structure in natural images. This thesis is about the hypothesis that the visual system relies on a good representation of images for its analysis. Here, good means such a representation which makes dependencies between pixel values explicit and easily accessible. May be as we have done it in Figure 1.4 by grouping locally those picture elements together which have dependencies. We will investigate this hypothesis by

“analysis by synthesis”, i.e. we aim at constructing such a representation of natural images and relate its properties to our visual system.

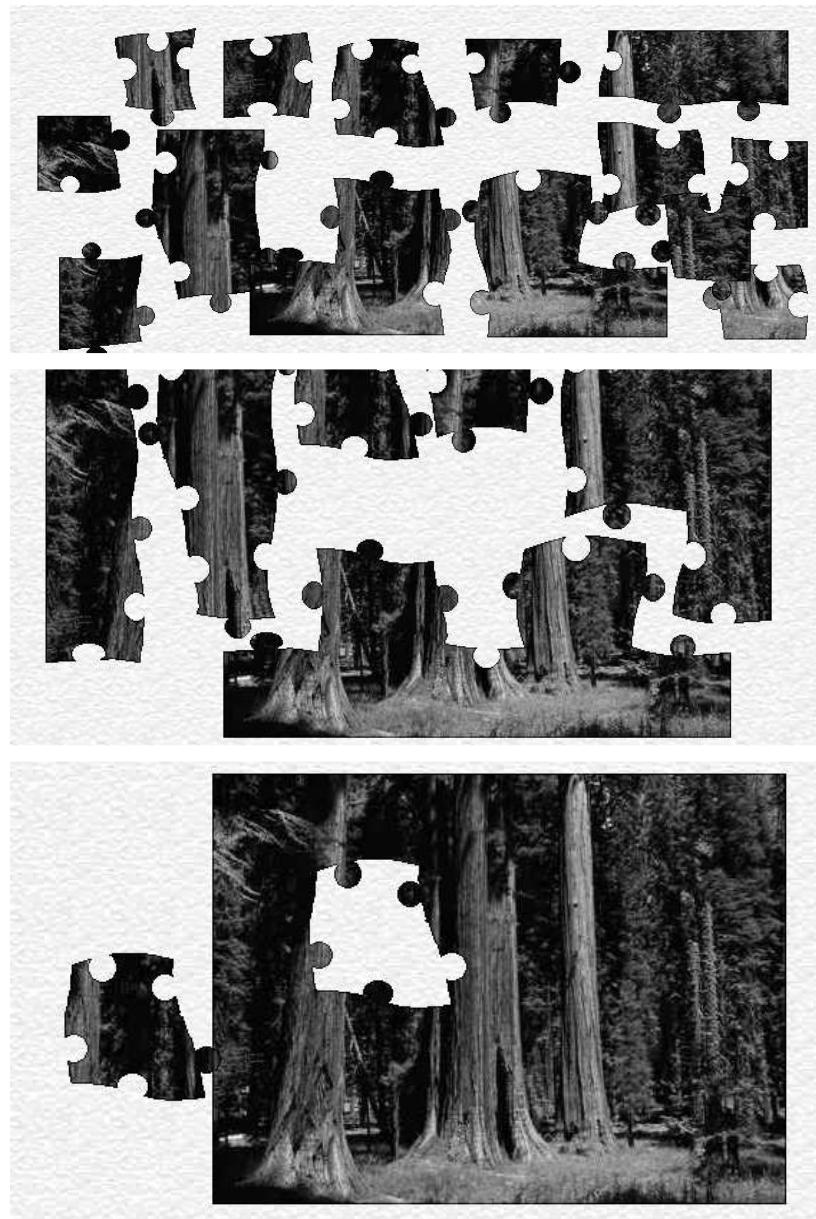


Figure 1.4: Discovering structure in the natural image by inside-out construction. Group picture elements locally together which show some similarities, then, leverage on that representation to yield iteratively groups of higher-order picture elements till the meaning of the image is discovered.

This thesis is structured as follows.

- In *Chapter 2*, we will explain the general background which is involved in the analysis of natural images and our visual system. We explain a mathematical concept which captures the notion of structure in images and present methods by which this structure can be made explicit. We give statistical properties of natural images and introduce briefly the mammal visual system. We close the chapter with information on engineering approaches to image analysis.
- In *Chapter 3*, we present previous studies which have investigated the aforementioned hypothesis that the visual system relies on a suitable representation of images to accomplish discovery of structure. We show how this hypothesis can account for properties of early stages of the visual system and present an approach which aims at testing the hypothesis on a higher-order area. We close the chapter with the formulation of our research questions.
- In *Chapter 4 to Chapter 6*, we present results of our investigations. In Chapter 4, we concentrate on image representations which involve the analysis of structure contained in one spatial scale only. In Chapter 5, we extend the investigations to multiple spatial scales. Chapter 6 is of more technical nature. We take a step back and scrutinize our methods to find ways to improve our investigations into the structure inherent in images.
- Conclusions from our research are drawn in *Chapter 7*.

Chapter 2

Background

2.1 Information theory

This section is a brief introduction to information theory. We define entropy, relative entropy, mutual information and redundancy. The notion of hidden redundancy is introduced and related to structure in random vectors.

2.1.1 Entropy

Entropy H is the basic concept of information theory, see [8]. It is defined for a continuous random vector \mathbf{x} with probability density p by the expectation of $\log \frac{1}{p(\mathbf{x})}$. Thus

$$H(\mathbf{x}) = \mathbb{E} \left[\log_2 \frac{1}{p(\mathbf{x})} \right]. \quad (2.1)$$

Entropy H can be interpreted as measure of uncertainty of the random vector \mathbf{x} and it is related to the minimum coding length needed to represent \mathbf{x} .

Entropy H is a functional of the distribution of \mathbf{x} . In turn, constraints on \mathbf{x} , such as a given value of expectation or variance, define a distribution which has maximum entropy H_{\max} given the constraints. For example, the uniform density has maximum entropy among the densities of given range. A Gaussian is the maximum entropy distribution for random variables of unit variance.

Denote by x_i a component of the random vector \mathbf{x} . Entropy is then bounded above by

$$H(\mathbf{x}) \leq \sum_i H(x_i) \quad (2.2)$$

with equality if and only if the x_i are independent.

2.1.2 Mutual information

Mutual information $I(\mathbf{x}_1, \mathbf{x}_2)$ is a measure of the amount of information that random vector \mathbf{x}_1 contains about random vector \mathbf{x}_2 [8]. It is defined by

$$I(\mathbf{x}_1, \mathbf{x}_2) = E_{p(\mathbf{x}_1, \mathbf{x}_2)} \left[\log_2 \frac{p(\mathbf{x}_1, \mathbf{x}_2)}{p(\mathbf{x}_1)p(\mathbf{x}_2)} \right]. \quad (2.3)$$

It is shown in [8] that

$$I(\mathbf{x}_1, \mathbf{x}_2) = H(\mathbf{x}_1) + H(\mathbf{x}_2) - H(\mathbf{x}) \quad (2.4)$$

for \mathbf{x} being the random vector $[\mathbf{x}_1 ; \mathbf{x}_2]^T$. Mutual information has the important property of being always nonnegative, and zero if and only if the involved variables are independent.

2.1.3 Redundancy

We define redundancy $R(\mathbf{x})$ of \mathbf{x} based on [35] by

$$R(\mathbf{x}) = 1 - \frac{H(\mathbf{x})}{H_{\max}(\mathbf{x})}. \quad (2.5)$$

As reported in [35], redundancy has been reformulated in [1] to emphasize two different origins

$$R(\mathbf{x}) = \frac{1}{H_{\max}} \left(H_{\max} - \sum_i H(x_i) \right) + \frac{1}{H_{\max}} \left(\sum_i H(x_i) - H(\mathbf{x}) \right). \quad (2.6)$$

The first term on the right hand side concerns the distribution of the individual components x_i of \mathbf{x} , irrespective of their dependencies. It measures how much the distribution of the individual x_i differs from the maximum entropy distribution. Consider the example when the range of the x_i is restricted. Then, the first term on the right hand side stands for redundancy due to non-uniformity of the x_i .

The second term on the right hand side has been defined in (2.4) as mutual information between the x_i . It describes redundancy due to statistical dependencies. It concerns the structure of the random vector \mathbf{x} . Figure 2.1 illustrates the differences between the two terms.

The latter term is called hidden redundancy whereas the former might be called obvious redundancy [3]. The names derive from the difficulty to discover hidden redundancy.

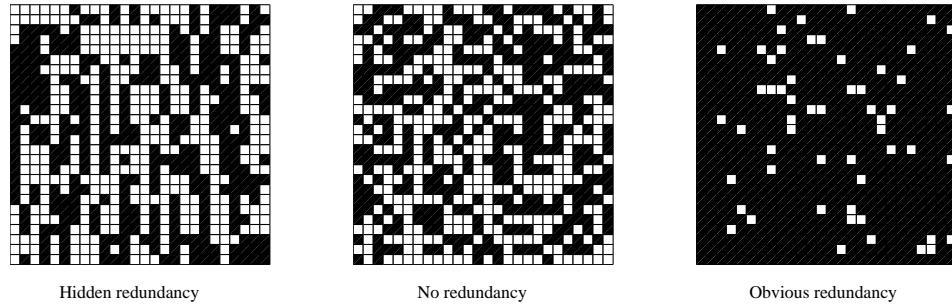


Figure 2.1: Illustration of the different origins of redundancy. All three images have been artificially constructed. Left: Sample image derived from the probability distribution $P(\mathbf{x}) = \prod_k \prod_j P(x_{3k+1,j}, x_{3k+2,j}, x_{3k+3,j})$ where x_{ij} denotes the pixel in row i , column j . See Table 2.1.3 for the definition of $p(x_{3k+1,j}, x_{3k+2,j}, x_{3k+3,j})$. Each pixel is Bernoulli distributed with success probability of one half. This is the maximum entropy distribution for a binary pixel. Redundancy is due to the vertical structure in the image, $R = 0.19$. Middle: Sample image with pixel values which are independent and uniformly distributed. Redundancy $R = 0$. Right: Sample image with pixel values which are independent and follow the density $P(\text{black}) = 0.9$, $P(\text{white}) = 0.1$. No hidden redundancy, but $R = 0.53$.

$x_{3k+1,j}$	$x_{3k+2,j}$	$x_{3k+3,j}$	P
0	0	0	$\frac{1}{3}$
0	1	0	$\frac{1}{30}$
0	0	1	$\frac{1}{15}$
0	1	1	$\frac{1}{15}$
1	0	0	$\frac{1}{15}$
1	1	0	$\frac{1}{15}$
1	0	1	$\frac{1}{30}$
1	1	1	$\frac{1}{3}$

Table 2.1: Probability $P(x_{3k+1,j}, x_{3k+2,j}, x_{3k+3,j})$ for the synthesis of the image on the left hand side in Figure 2.1. The three vertical pixels have dependencies which lead to images with vertical structure. The marginal distribution of each pixel is a Bernoulli distribution with success probability of one half.

2.2 Independent component analysis

In this section, we define linear independent component analysis by an information-theoretic approach. We highlight non-Gaussianity of the independent components and relate independent component analysis to hidden redundancy reduction. We introduce the FastICA algorithm and explain the origin of the nonlinearity which drives the learning process.

2.2.1 Purpose

Linear independent component analysis (ICA) is a statistical method to find a suitable representation of multidimensional data, i.e. realizations of the random vector \mathbf{x} . It looks for a linear transformation \mathbf{W} which yields components s_i that are non-Gaussian and statistically as independent from each other as possible [16] [17] [21]. The components s_i are considered the hidden reasons which caused the data.

2.2.2 Information theoretic approach

The ICA of a random vector \mathbf{x} is defined in [7] [16] as an invertible linear transformation

$$\mathbf{s} = \mathbf{W}\mathbf{x}, \quad (2.7)$$

where the deterministic weight matrix is \mathbf{W} determined such that mutual information between the components s_i is minimized. Since mutual information is invariant to scaling of the involved random variables, the independent components s_i are defined up to multiplicative constants. With reference to (2.6), the random vector \mathbf{x} is in ICA linearly transformed such that it contains less hidden redundancy. However, this does not imply that overall redundancy is reduced either.

Mutual information of the s_i is shown in [16] to be

$$I(\mathbf{s}) = \sum_i H(s_i) - H(\mathbf{x}) - \log |\det \mathbf{W}|. \quad (2.8)$$

Let the s_i be uncorrelated and of unit variance. In that case, the s_i become defined up to their sign. Moreover, the determinant of \mathbf{W} is constant and minimizing (2.8) with respect to \mathbf{W} is equivalent to minimizing $\sum_i H(s_i)$. We have noted in Subsection 2.1.1 that the Gaussian distribution achieves maximum entropy for variables of unit variance. Since the s_i are

constrained to unit variance, the latter minimization can be interpreted as sum of non-Gaussianities [21]. This highlights a fundamental guiding principle in ICA, i.e. “Independent components are the maximally non-Gaussian components” [21].

With the definition of entropy (2.1), the optimization problem becomes

$$\mathbf{W} = \arg \max_{\tilde{\mathbf{W}}} \sum_i E [\log p_i(\tilde{\mathbf{w}}_i^T \mathbf{x})], \quad (2.9)$$

where the vector \mathbf{w}_i^T denotes a row of the matrix \mathbf{W} . The matrix \mathbf{W} is a functional of the probability densities p_i . They are normally not known and have to be estimated. This leads to an approximation of mutual information and establishes a connection to ICA as maximum likelihood estimation of a linear mixing model [21].

The inverse of the weight matrix \mathbf{W} is called the mixing matrix \mathbf{A} and the mixing model is then given by

$$\mathbf{x} = \mathbf{As}. \quad (2.10)$$

This means, ICA defines a generative model for the random vector \mathbf{x} . The random vector is assumed to be a linear mixture of mutually independent and non-Gaussian random variables s_i . The mixing system is not assumed to be known. ICA estimates the mixing model, i.e. the mixing matrix \mathbf{A} and the components s_i , the hidden reasons for \mathbf{x} .

2.2.3 Fixed-point algorithm

We state a robust and fast algorithm, the FastICA algorithm [16] [17] [21] for one unit. It finds row \mathbf{w}_i^T of the weight matrix \mathbf{W} in (2.7) by optimization of one summand in (2.9). The algorithm finds a local maximum of

$$E [G(\mathbf{w}_i^T \mathbf{x})] \quad (2.11)$$

with an approximative Newton’s method. The nonlinear function G can then be interpreted as the logarithm of a probability density p_i or as a measure of non-Gaussianity. Its derivative G' is denoted by g .

Denote the whitening matrix of the data \mathbf{x} by \mathbf{V} , then the basic FastICA iteration for the white, zero mean data $\mathbf{z} = \mathbf{Vx}$, is given by

$$\mathbf{b}_i \leftarrow E [\mathbf{z}g(\mathbf{b}_i^T \mathbf{z})] - E [g'(\mathbf{b}_i^T \mathbf{z})] \mathbf{b}_i \quad (2.12)$$

$$\mathbf{b}_i \leftarrow \mathbf{b}_i / \| \mathbf{b}_i \| . \quad (2.13)$$

The column vector \mathbf{w}_i which maximizes (2.11) is found from \mathbf{b}_i with

$$\mathbf{w}_i = \mathbf{b}_i^T \mathbf{V}. \quad (2.14)$$

The FastICA algorithm for parallel estimation of all the column vectors of the matrix \mathbf{W} is presented in [16] [17] [21] including a discussion of its properties.

2.3 Natural images

We explain in this section that natural images are highly structured and that this causes strong dependencies between different picture elements. We present the idea of scale invariance and point out why second-order statistics are not enough to capture the structure in natural images.

2.3.1 Structure in natural images

Consider a discrete (quantized) image $I(x, y)$ which is made of N picture elements (pixels). Moreover, assume that each pixel can represent M different luminance values. Let us furthermore write the image $I(x, y)$ column-wise as vector \mathbf{x}^I of dimension N . Then, the image vector \mathbf{x}^I can represent $M \cdot N$ different images. However, an inspection of all these images reveals that only a small sub-manifold corresponds to natural images, i.e. wildlife scenes as landscapes our ancestors might have seen. That is, natural images are a small subset in the image space [31]. In other words, natural images are highly structured.

Given the large number $M \cdot N$ of possible images, it is helpful to use information and probability theory to characterize natural images. A natural image is seen as a realization of a random vector \mathbf{x}^I . Because of the structure in natural images, \mathbf{x}^I is non-uniformly distributed and importantly, its pixels have dependencies among each other. With (2.6), this can be paraphrased by saying that because of the structure in natural images, \mathbf{x}^I contains hidden redundancies.

2.3.2 Scale invariance

A common assumption is that image probabilities are invariant to translation [34]. The autocorrelation function becomes then a function of the

distance between pixels only. Its Fourier Transform, i.e. the power spectral density of \mathbf{x} , is reported in [33] to fall like $1/f^2$, where f is the spatial frequency.

One of the most commonly interpretation of this power law is that it is due to scale invariance of the visual world [33]. Scale invariance means that the statistical properties of images should not change if the scale of observation is changed. Power contained in a spatial frequency band had then to be the same regardless of scale. A Fourier spectrum must fall like $1/f^2$ to comply with that constraint.

2.3.3 Non-Gaussianity

The probability distribution of \mathbf{x}^I is believed to be non-Gaussian. Direct evidence is given by the finding that response distributions of linear oriented bandpass filters to natural images have sharp peaks at zero, and much longer tails than a Gaussian density. Such a distribution is called a super-Gaussian, or sparse distribution, see Figure 2.2. Since any marginal density of a Gaussian has a Gaussian distribution, this constitutes direct evidence for non-Gaussianity of natural images [33].

Furthermore, if natural images were Gaussian all the statistical information should be contained in the power spectral density. However, images that are sampled by filtering white noise with a filter which has a $1/f$ transfer function are devoid of any edges, contours or other structures found in natural images. Similarly, if natural images were Gaussian whitening should yield independent Gaussian coefficients [33]. However, a whitened natural image is still very structured. See Figure 2.2. This is so because localized structures in natural images are characterized by their Fourier phase spectrum. Edges for example, have local phase alignments in spatial frequency [29]. Whitening, however, considers only linear correlation, i.e. the power spectral density, and is thus blind to Fourier phase alignment.

2.4 Primary visual cortex

We present in this section very briefly the pathway from the eye to the primary visual cortex. We explain how cortical cells differ from cells earlier in the pathway. Mathematical models of cortical cells are presented and functional interpretations of cortical cells are given.

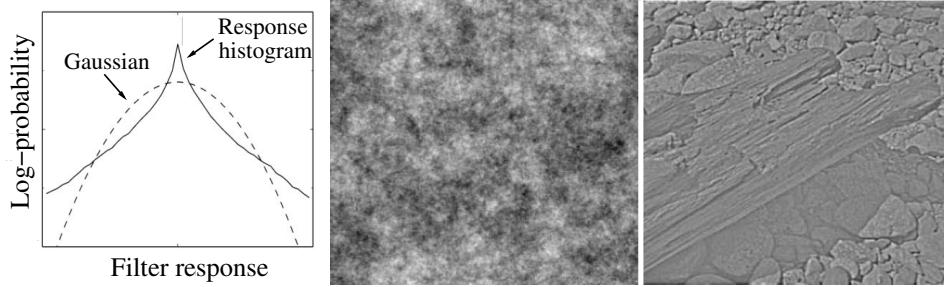


Figure 2.2: Non-Gaussianity of natural images. Left: Histogram of responses of an oriented bandpass filter to a natural image. Middle: Sample of $1/f$ white noise. Right: Whitened natural image. Adapted from [33].

2.4.1 Visual pathway

The eyes are the front end of the visual system. They gather light reflected from objects in the environment and focus it on the retina. The retina is covered with photo-receptors which convert light into neural signals which are carried via the optic nerve to the Lateral Geniculate Nucleus (LGN) and then to the primary visual cortex (V1) at the back of the brain. See Figure 2.3.

Photo-receptors have a logarithmic response to light and adapt to lightening conditions. Retinal and geniculate cells have center-surround receptive fields which means that the cells compare the amount of light hitting a certain spot on the retina with the average amount falling on the immediate surround. Diffuse light is ignored. See Figure 2.4(a) for an illustration and [14] for a discussion of the neural circuitry prior to V1.

2.4.2 Models of cortical cells

Responses of cortical cells to visual stimuli summarize all processing from the retina till V1. Unlike the receptive fields of cells prior to V1 with their circular symmetry, cortical cells have oriented receptive fields, see Figure 2.4(b). Neurons in V1 have been classified in three categories [14]: simple-cells, complex-cells and end-stopped cells. We present their response properties to static stimuli and indicate their mathematical modeling.

Basis of the presented models is a linear model. Denote the visual stimulus by $I(x, y)$. In order to incorporate adaptation and logarithmic response

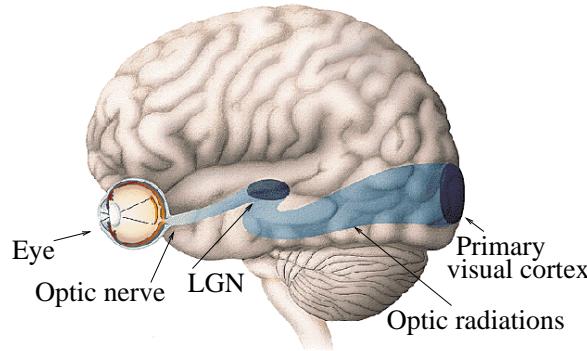


Figure 2.3: Visual pathway from the eye to the primary visual cortex. Responses of cortical cells to visual stimuli include processing in stages earlier to V1.

properties on the retinal level, $I(x, y)$ is taken as a quantity which is proportional to changes from average image intensity, i.e. contrast. The linear model is then as follows

$$s = \langle w(x, y), I(x, y) \rangle, \quad (2.15)$$

where $\langle \cdot \rangle$ denotes the inner product. Output s is interpreted as deviation of the firing rate from the average firing rate of the neuron. Positive values indicate that the input has excitatory effect on the neuron and negative values indicate an inhibitory effect. The feed-forward weight $w(x, y)$ is interpreted as the receptive field of the neuron.

It is practical to vectorize the above quantities. The feed-forward weight $w(x, y)$ and the image $I(x, y)$ are organized as column vectors \mathbf{w} and \mathbf{x}^I . The linear neuron model can then be written as

$$s = \mathbf{w}^T \mathbf{x}^I. \quad (2.16)$$

Simple-cells

A simple-cell S is classically modeled as linear oriented bandpass filter [6] [22]. The Gabor receptive field model is

$$w_G(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp \left[-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2} \right] \cos(2\pi f x - \phi), \quad (2.17)$$

where the coordinates (x, y) are given by the coefficients of the rotated basis vectors \mathbf{e}_x and \mathbf{e}_y in Figure 2.5.

The Gabor filter in Figure 2.5 is tuned to frequency f and orientation θ . The following quantities are further used to characterize a simple-cell [9] [10]:

- Length l_y is defined as the full width at half maximum of the envelope along the orientation into which the filter is low-pass. For the Gabor receptive field model,

$$l_y = 2\sqrt{2 \ln(2)}\sigma_y. \quad (2.18)$$

- Width l_x is defined as the full width at half maximum of the envelope along the orientation into which the filter is bandpass. For the Gabor receptive field model,

$$l_x = 2\sqrt{2 \ln(2)}\sigma_x. \quad (2.19)$$

- Aspect ratio l is the ratio of length and width,

$$l = \frac{\sigma_y}{\sigma_x}. \quad (2.20)$$

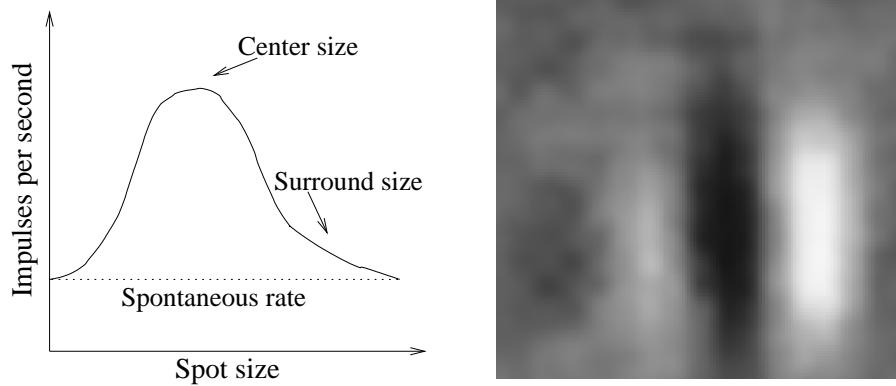
- Frequency bandwidth Δ_f of the filter is defined as the difference between the highest and the lowest frequency that yield half of the maximal response in octaves. For the Gabor receptive field model,

$$\Delta_f = \log_2 \left[\frac{2\pi\sigma_x f + \sqrt{2 \ln(2)}}{2\pi\sigma_x f - \sqrt{2 \ln(2)}} \right]. \quad (2.21)$$

- Orientation bandwidth Δ_θ is defined as the difference in orientation from the peak orientation θ for which the response has declined to half of its peak value. For the Gabor receptive field model,

$$\Delta_\theta = 2 \sin^{-1} \left[\frac{1}{l} \cdot \frac{2^{\Delta_f} - 1}{2^{\Delta_f} + 1} \right]. \quad (2.22)$$

Bandwidth Δ_f of simple-cells in V1 is observed to be fairly constant, irrespective of peak frequency tuning f [9]. A typical value is $\Delta_f = 1.5$ oct [9]. Aspect ratio l is in the range $l = 4$ to $l = 1$, typically around $l = 1.5$ [22]. The spatial phase ϕ ranges uniformly in the range 0–90 degree [9]. Frequency tuning is observed to fall along a continuum [31].



(A) Stimulation of a retinal ganglion cell with ever larger spots increases the response, up to a spot size where further enlargement causes a decline. The cell has a center-surround receptive field. Adapted from [14].

Figure 2.4: The receptive field is the area on the retina from where a neuron can be influenced. Unlike the receptive fields of cells prior to V1 with their circular symmetry, cortical cells have oriented receptive fields.

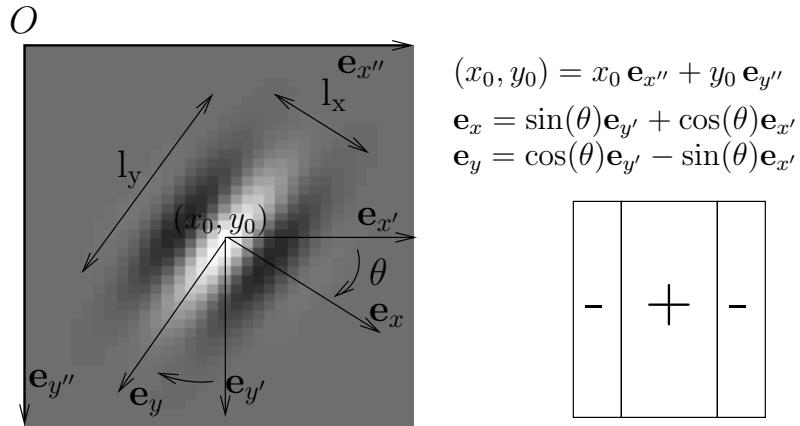


Figure 2.5: Gabor receptive field model as defined in (2.17) and abstraction. Bright shading indicates a high value of $w(x, y)$ and dark shading a low value. The Gabor filter is centered at (x_0, y_0) and oriented to θ . It is of length l_y and width l_x . The filter is low-pass along the axis \mathbf{e}_y and bandpass along the axis \mathbf{e}_x .

Complex-cells

A complex-cell \mathcal{C} shares with a simple-cell S the property that it can be modeled as an oriented bandpass filter. complex-cells show however nonlinear behavior. One aspect is their relative invariance to phase of the stimulus. Classically, they are modeled with the energy model as a quadrature phase filter [9]. The response of a complex-cell \mathcal{C} to the stimulus $I(x, y)$ is given by

$$\mathcal{C}[I] = \langle w_e(x, y), I(x, y) \rangle^2 + \langle w_o(x, y), I(x, y) \rangle^2. \quad (2.23)$$

The filter $w_e(x, y)$ is an even Gabor filter w_G , i.e. $\phi = 0$ in (2.17) and $w_o(x, y)$ is an odd Gabor filter, i.e. $\phi = \pm\pi/2$ where other parameters are the same.

End-stopped cells

An ordinary simple or complex-cell usually shows length summation. The longer the stimulus, the better is the response, until the elongated stimulus is as long as the receptive field. Making it still longer has no effect any more. For an end-stopped simple or complex-cell, lengthening the stimulus improves the response up to some limit, but exceeding that limit results in a weaker response. For completely end-stopped cells response would completely vanish [14].

One explanation is that several cells of the same orientation converge on the end-stopped cell. Those, whose receptive field are congruent with the end-stopped cell's activating region make excitatory contacts. The others make inhibitory contacts [14].

An alternative explanation involves convergence of cells with receptive fields of different sizes. A inhibiting cell covers the entire area and responds only weakly to short oriented stimuli. If a excitatory cell of smaller receptive field responds strongly to the same stimuli, the cell onto which both the inhibiting and excitatory cell converge would become end-stopped [14].

2.4.3 Function of cortical cells

There are various propositions about the function of cortical cells. One hypothesis of what the cells are doing is that they act as line and edge detectors. This idea is motivated by experimental evidence that cortical cells do not strongly respond to spots of light, like retinal or geniculate cells, but to stimuli with elongated intensity changes, like edges or bars.

An alternative functional interpretation of cortical cells derives from local spatial frequency theory. It is suggested that cortical cells are not detectors of image features, as lines or edges, but are general purpose filters which perform a local Fourier analysis of the stimulus. According to this theory, line and edge detection would be located at a higher level than V1. More suggestions on the function of cortical cells are discussed in [31].

2.5 Edge detection

We explain in this section the principles and difficulties in edge detection. We present design criteria for linear filters and emphasize the need of filters of different scales. Advantages of quadratic filtering are given.

2.5.1 Difficulties in edge detection

Edge detection concerns the localization of significant variations in intensity of the image. An overview of the variety of detection techniques is found in [38]. Edge detectors are commonly such that they produce an edge map as output where local maxima indicate an edge in the original image. Main concern is that a local maxima is a reliable indicator of the presence of an edge. That is, a single edge in the original image should be unambiguously linked to a local maximum. Noise in the image and uncertainties about what “significant variation” actually means makes this difficult however.

2.5.2 Linear filtering approach

The edge map is constructed by convolution of the image $I(x, y)$ with a filter $w(x, y)$. The linear filtering operation corresponds to smoothing and the calculation of the derivative. Smoothing is to reduce noise, differentiation to detect the significant variations in the image [38]. The basic idea of the linear filtering approach is thus to look for maxima of the gradient of the image.

Design criteria to detect one dimensional step edges in white Gaussian noise have been proposed [5]. For an image $I(x, y)$ which has an one dimensional step edge at x_0 , the filter $w(x)$ should be such that [5]:

- The maximum of the output $s(x)$ of the convolution occurs close to x_0 , even under noisy conditions. This is the localization criterion.

- The signal to noise ratio is high at x_0 , that is $s(x_0)$ takes low values for white Gaussian noise input, but high values for an edge stimulus. This is the detection criterion.
- There are no local maxima of $s(x)$ near x_0 which would cause ambiguities about the precise location of the edge. This is the multiple response criterion.

The derivative of a Gaussian has been found to be a filter $w(x, y)$ that optimizes approximately the above criteria [5]. Gabor filters have also been designed to optimize them [26].

It is shown in [5] that detection performance improves with the broadness of the filter, e.g. the Gaussian. Localization performance however decreases. For a filter of given broadness, or scale, localization and detection are therefore conflicting. The need of multiple scales to resolve that problem has been pointed out in [5].

In order to detect a two dimensional edge of orientation θ , it has been proposed to enhance the optimal filter with a Gaussian window function along the y -axis and to use this combined filter rotated by θ to detect the edge. It is shown in [5] that both localization and detection performance improves with the aspect ratio of the two dimensional filter.

It has become clear that a combination of filters of multiple scales and orientations should be involved in edge detection in two dimensions. The effective way of how they should be combined has albeit proven to be a problem on its own.

2.5.3 Quadratic filtering approach

In the formulation of the design criteria for linear filters, significant variation has been interpreted as a step edge. This is however difficult to justify. Figure 2.6 shows that interesting variations in image intensity can also be caused by composite edges, i.e. edges which are a combination of lines, steps and roofs.

It has been shown in [32] that for any linear filter, there is always a composite edge such that the location of the maximum of the filtered image does not correspond to the location of the edge. Moreover, it has been shown that arbitrary precise localization of composite edges can be achieved by looking for peaks in the response to a quadratic filter, i.e. a filter of the

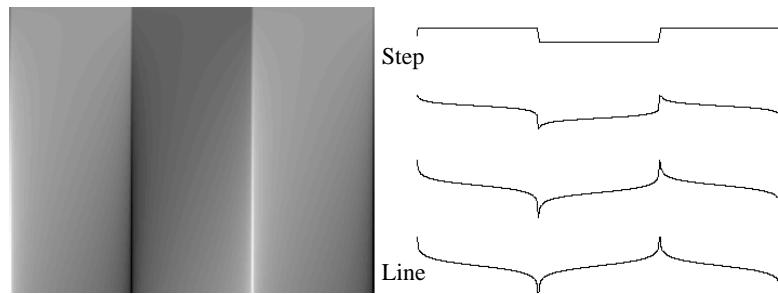


Figure 2.6: A composite edge. Combinations of steps, lines and roofs are all edges. Left: image. Right: Horizontal slice trough image intensity. Adapted from [25].

form $\sum_i \langle w_i(x, y), I(x, y) \rangle^2$ with suitably chosen linear filters $w_i(x, y)$. These filters might involve multiple scales to improve performance.

A special case of quadratic filters are quadrature phase filters where two filters are used which form a Hilbert pair. The complex cell model (2.23) is a quadrature phase filter. An idea behind quadrature phase filters is that interesting variations in image intensity are caused by Fourier phase congruency of the image [2].

Chapter 3

Previous Work and Research Questions

3.1 Learning properties of the visual cortex

In this section, we explain the research hypothesis which underlies the previous work to this thesis and the thesis itself. We present how it is tested by “analysis by synthesis”.

3.1.1 Theoretical framework

The visual cortex forms a complete system which performs well at pattern analysis - the problem of extracting structure out of data. Although there are ideas about the function of parts of the visual cortex, e.g. edge detection for simple-cells as described in Subsection 2.4.2, the principles of its operation are not well understood [27].

The structure of natural images, see Subsection 2.3.1, is a starting point to understand the principles of operation. The idea is that statistical properties of natural images, i.e. wildlife scenes, have fundamentally influenced the design of the visual cortex by either forces of evolution or adaptation. Since the pixel-by-pixel representation of natural images contains a lot of redundancies, see Subsection 2.3.1, the visual cortex must have discovered means to represent them in a more suitable way [29] [30]. Work prior to this thesis as well as the thesis itself is based on a research hypothesis which belongs to the theoretical framework of optimal encoding [37]. The hypothesis is that a suitable representation for the visual cortex is such a representation which

makes the structure in natural images explicit. In other words, a representation with as little hidden redundancy as possible [3]. See Subsection 2.1.3 for a discussion of hidden redundancy.

Motivation for such a representation is given in [3] [27] from the viewpoint of perception as probabilistic inference [24] [36]. Moreover, it has been pointed out in [12] that reducing hidden redundancy implies discovering the structure of an image - the job of any visual system.

3.1.2 Analysis by synthesis

The research hypothesis of hidden redundancy reduction is tested with analysis by synthesis.

- A system to represent natural images is constructed with the research hypothesis as design rule. That is, a statistical model for natural images is learned such that hidden redundancies are pinpointed.
- The obtained representation is related to properties of the visual cortex to test its predictive and explanatory power. Have any properties of the visual cortex been learned?
- If this is the case, the design rule which has driven the learning process, i.e. hidden redundancy reduction, might be related to the principles of patterns analysis employed by the visual cortex.

3.2 Learning simple-cell receptive fields

Denote by \mathbf{x}^I the random vector which is a vectorized natural gray-scale image as in Subsection 2.3.1. The image \mathbf{x}^I can be written as linear superposition

$$\mathbf{x}^I = \sum_i \mathbf{a}_i^S s_i^S. \quad (3.1)$$

The point is that this new representation of an image by means of the deterministic basis vectors, or generative weights, \mathbf{a}_i^S and the random variables s_i^S should make the structure in the image \mathbf{x}^I more explicit. This is mathematically implemented by demanding the s_i^S ,

$$s_i^S = \langle \mathbf{w}_i^S, \mathbf{x}^I \rangle, \quad (3.2)$$

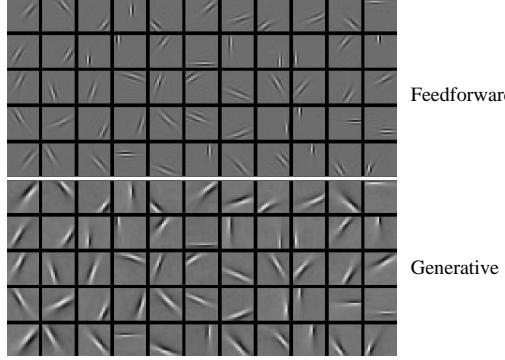


Figure 3.1: Generative and feed-forward weights learned with ICA from natural images. They share many properties of simple-cells, e.g. they are localized, oriented bandpass filters. The generative weights are a low-pass-filtered version of the feed-forward weights. Adapted from [10].

to be maximally independent and to follow a super-Gaussian, or sparse, distribution. Super-Gaussian random variables take more often either small or large absolute values than a Gaussian random variable, see Figure 2.2. This means that any image, i.e. any realization of the random vector \mathbf{x}^I , can be described with only a few non-zero coefficients given by the realization of the random variables s_i^S . The meaning of the coefficient s_i^S is given in terms of images by the generative weight \mathbf{a}_i^S .

This image representation is called sparse coding or with reference to Section 2.2 ICA with a sparse prior. When learned from natural images, it leads to generative weights \mathbf{a}_i^S and feed-forward weights \mathbf{w}_i^S which show simple-cell receptive field properties [4] [10] [28] [30]. See Figure 3.1.

ICA is interpreted in Section 2.2 as a method to make hidden redundancies manifest. This suggests that simple-cells are part of the operation-principle of the visual cortex to represent the structure in images in a explicit manner. Furthermore, it suggests that properties of simple cells S_i , see e.g. Subsections 2.4.2 and 2.4.3, are determined by the statistics of natural images.

3.3 Learning complex-cell receptive fields

There is structure in the s_i^S , hence also in \mathbf{x}^I , which is not made explicit in the representation in (3.1). In a modified version of ICA, in Independent

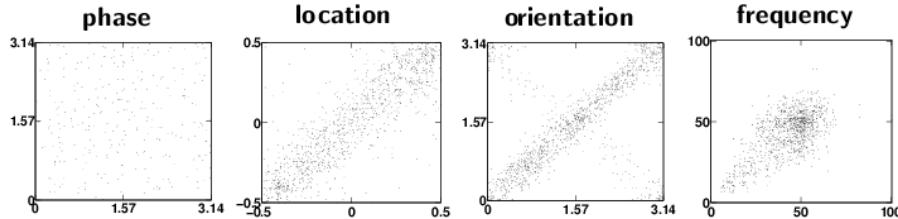


Figure 3.2: Local parameter correlation of filters in a common complex cell subspace. Each pair in a filter subspace is one dot and axes give the parameter values for fitted Gabor filters. From [19].

Subspace Analysis (ISA), filters \mathbf{w}_{ij}^C are sought which make s_i^C ,

$$s_i^C = \sum_j \langle \mathbf{w}_{ij}^C, \mathbf{x}^I \rangle^2, \quad (3.3)$$

as independent as possible and follow a sparse distribution [18].

When learned from natural images, the filters \mathbf{w}_{ij}^C become localized oriented bandpass filters as the \mathbf{w}_i^S . Particular about the filters $\mathbf{w}_{i_0j}^C$, for a fixed $i = i_0$, is that they share localization, orientation and frequency tuning properties, not however phase properties. See Figure 3.2. That is, the subspace i_0 spanned by the filters $\mathbf{w}_{i_0j}^C$ defines approximately a quadrature phase filter. The norm of the projection of an image I on that subspace is given by $s_{i_0}^C$. It is interpreted as the response $\mathcal{C}_{i_0}[I]$ of a complex-cell \mathcal{C}_{i_0} to the stimulus I .

Additional structure in natural images is made explicit by the subspaces. All the filters \mathbf{w}_{ij}^C which produce outputs that are correlated in energy are grouped into a single subspace i [19] [20]. Whereas correlation in energy means that the square rectified outputs are correlated not however the outputs themselves.

Both complex-cell response behavior and simple-cell response behavior can thus be understood as being part of a single image processing strategy of the visual cortex, namely making hidden redundancies in static gray-scale images explicit. Strategies which include the processing of time dependent images are addressed in [15] [23].

3.4 Learning cells beyond V1

A way to go beyond V1 is to seek a good representation of the complex-cell outputs s_i^C . In order to explore structure in the s_i^C , complex-cell responses $\mathcal{C}_i[I]$ have been modeled in [13] as outputs x_i^C of quadrature phase filters as in (2.23), i.e. in vectorized notation

$$x_i^C = (\mathbf{w}_{e_i}^T \mathbf{x}^I)^2 + (\mathbf{w}_{o_i}^T \mathbf{x}^I)^2, \quad (3.4)$$

with \mathbf{w}_{e_i} a vectorized even Gabor filter $w_e(x, y)$ and \mathbf{w}_{o_i} an odd Gabor filter. Only Gabor filters tuned to one spatial frequency f were used.

In line with Section 3.2, the outputs x_i^C are described as a linear combination of outputs s_i^H of higher-order cells \mathcal{H}_i ,

$$\mathbf{x}^C = \sum_i \mathbf{a}_i^H s_i^H. \quad (3.5)$$

This model is estimated in [13] with non-negative sparse coding [11] from natural images. Generative weights \mathbf{a}_i^H are sought such that the s_i^H become as independent and sparse as possible under the constraint that all the components of \mathbf{a}_i^H and the s_i^H are non-negative.

It is shown in [13] that the learned generative weights \mathbf{a}_i^H code the simultaneous activity of collinear complex-cells. See Figure 3.3. Collinear activity of complex-cells, i.e. a complex-cell contour, is seen as an indicator for a contour in an image. This means that the strategy of making structure in natural images explicit would account for contour-coding cells.

3.5 Research questions

Now, we motivate and formulate the research questions of this thesis. They are based on previous work described in Section 3.4. They address the need of *non-negative* sparse coding to learn cells beyond V1, the case where multiple scales are allowed and the validity of the original experimental design.

3.5.1 Standard ICA model of complex-cell outputs

In previous work described in Section 3.4, the statistics of complex-cell outputs are modeled by means of non-negative sparse coding. This has led to the emergence of contour cells.

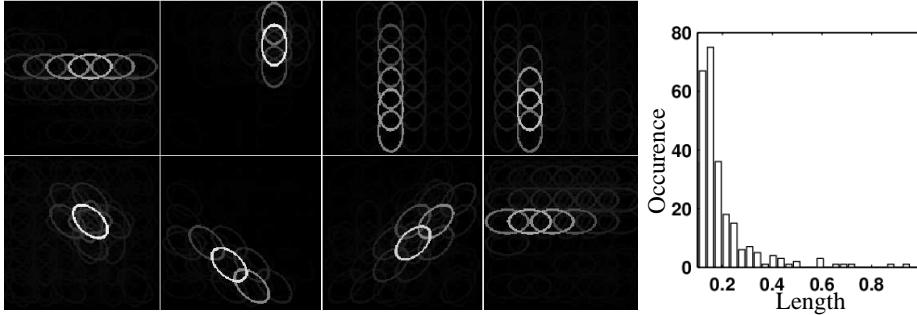


Figure 3.3: Left: Higher-order cells coding the simultaneous activity of collinear complex-cells, i.e. complex-cell contours. Brightness indicates the strength of activity. Black means zero activity, bright means high activity. Right: Length distribution of the contours relative to the width of the sampling window. Adapted from [13].

We are interested in the necessity of the non-negativity constraint. Is non-negativity essential for contour cells to be learned?

Non-negative sparse coding without the non-negativity constraint burns down to ICA. That is, ICA implements hidden redundancy reduction in a more parsimonious way than non-negative sparse coding. If contour cells emerged for standard ICA, a single parsimonious learning algorithm would synthesize image representations similar to those found in V1 and beyond. This would strengthen the evidence that making hidden redundancies explicit is related to the strategy employed by the visual cortex. That is, it would strengthen the research hypothesis of Subsection 3.1.1.

3.5.2 The case of multiple scales

In previous work described in Section 3.4, complex-cell responses were computed for filters which were all tuned to the same spatial frequency.

We are interested in structure contained in the outputs of complex-cells tuned to multiple frequencies. What kind of representation is learned by ICA for complex-cells which are not all tuned to the same spatial frequency?

Motivation for this is that by Fourier theory, multiple frequencies are needed to represent real-world structures. A higher-order feature consisting only of a single frequency is, in contrast, an artificial construct. Moreover, frequency tuning in V1 was found to fall along a continuum, see Subsection 2.4.2. Consequently, we are moving towards a richer model for V1 if we

consider complex-cells which are not all tuned to the same frequency.

3.5.3 Validity of the experimental design

In previous work described in Section 3.4, it was claimed that the learned higher-order cells make manifest hidden redundancies which are due to contours in natural images. The contours, however, reflect effectively structure in the data extracted with a filterbank, and not directly structure in natural images.

We are interested in the influence which a filterbank has on the representation of complex-cell outputs which is learned by ICA. How well adapted is the experimental design in Section 3.4 to the task of testing hidden redundancy reduction on areas beyond V1?

The motivation is to improve the way adopted for testing the research hypothesis of Subsection 3.1.1 on areas beyond V1.

Chapter 4

Standard ICA Model of Complex-Cell Outputs

4.1 Introduction

In this chapter, we present the use of standard ICA to reduce hidden redundancy in complex-cell responses to natural images as it has been announced in Subsection 3.5.1. Does standard ICA lead to complex-cell contours as non-negative sparse coding did?

We used handmade quadrature phase filters to emulate complex-cells according to the energy model. With these filters, we calculated complex-cell responses to natural images. These responses were analyzed by ICA to make structure in them explicit. We will show how this approach involves free parameters that had to be chosen. We will identify them and present to which extent complex-cell representations differed when they were learned with different parameter settings.

4.2 Methods

This section explains the methods which are applied in this chapter. Note that they are also basis of those of the following chapters. Figure 4.1 gives an overview of the different steps of the experiment. The steps are presented in detail in Subsection 4.2.1. In Subsection 4.2.2, we summarize free parameters and discuss their influence on the experiment. Methods which were adopted for interpretation and visualization of the learned ICA model are given in Subsection 4.2.3.

4.2.1 Experimental design

Data

Source images were 1008 gray-scale natural images of size 1024×1536 pixels from van Hateren's database¹ [10]. The images were natural images in the narrower sense, i.e. only wildlife scenes. From the source images, 50000 image patches were randomly extracted. An equal number was taken per image. The size of an image patch was 24×24 pixels. Mean gray value of each image patch was subtracted and the luminance values were rescaled to unit variance. The resulting image patch was thus proportional to changes from average image intensity. It shall be denoted by I and its vectorized version by \mathbf{x}^I .

Filtering

The filter stage was the same as in [13]. Statistics of the \mathbf{x}^I were extracted as shown in Figure 4.2. The filterbank consisted of complex-cells \mathcal{C}_i arranged on a 6×6 grid. Complex-cell responses to natural images $\mathcal{C}_{(x,y,\theta,f)}[\mathbf{x}^I]$ were modeled with the energy model (2.23). The coordinates (x, y) denote the center position of a complex-cell and correspond to the location of a node on the grid in Figure 4.2. Each node was occupied by complex-cells oriented to $\theta = k \cdot \frac{\pi}{4}$, $k = 0, \dots, 3$. Spatial frequency tuning was the same for all cells, $f = 0.22$ cycles/pixel. See Figure 4.2 for details.

Normalization

Each complex-cell output $\mathcal{C}_{(x,y,\theta,f)}[\mathbf{x}^I]$ was normalized by application of the function

$$f_\alpha(y) = y^\alpha, \quad (4.1)$$

or the function

$$f_\beta(y) = \begin{cases} y & \text{if } y < 1, \\ y^\beta & \text{if } y > 1. \end{cases} \quad (4.2)$$

When f_β was applied, it was assumed that the expectation of $\mathcal{C}_{(x,y,\theta,f)}[\mathbf{x}^I]$ was 1.

We shall denote by \mathbf{x}^C the vector which collects all the normalized complex-cell outputs. We numbered the complex-cells in a unique way such

¹Available at <http://hlab.phys.rug.nl/imlib/index.html>. Category “deblurred”.

that a complex-cell located at (x, y) , oriented to θ and tuned to f had number k ,

$$k = f_k(x, y, \theta, f), \quad (4.3)$$

for a bijective function f_k . The normalized response of complex-cell \mathcal{C}_k shall then be denoted by $\mathbf{x}^C(k)$.

Learning algorithm

As in [13], the (normalized) complex-cell responses \mathbf{x}^C were described as a linear combination of outputs s_i^H of higher-order cells \mathcal{H}_i . See Section 3.4 for a brief review of [13]. The mixing model

$$\mathbf{x}^C = \sum_i \mathbf{a}_i^H s_i^H \quad (4.4)$$

was estimated with the FastICA algorithm which was presented in Section 2.2. For that purpose, the complex-cell responses were first whitened, and possibly, the dimension was reduced by Principal Component Analysis (PCA). See [21] for a discussion of PCA and dimension reduction. Table 4.1 summarizes the different nonlinearities g which we have used with FastICA.

Post-processing

It has been pointed out in Section 2.2 that there is a sign ambiguity for the \mathbf{a}_i^H and the s_i^H . For each generative weight vector \mathbf{a}_i^H , the sign was chosen such that the sign of the component with the maximal absolute value was positive, i.e.

$$\text{signum} [\mathbf{a}_i^H(k^*)] > 0 \quad (4.5)$$

for

$$k^* = \arg \max_k |\mathbf{a}_i^H(k)|. \quad (4.6)$$

4.2.2 Free parameters

The Gabor filters which underlay the used complex-cells were the basic statistical analyzers of the images. Center location (x, y) , orientation θ , frequency tuning f , aspect ratio l and frequency bandwidth Δ_f had to be determined for each filter. This defined the filterbank that was used to extract the statistical properties of the images. The parameter were chosen

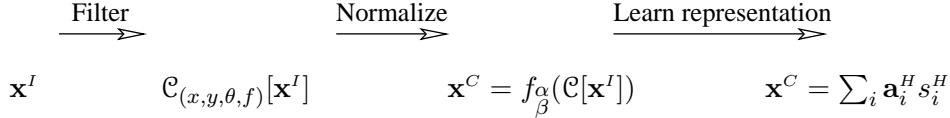


Figure 4.1: Methods overview. Image patches \mathbf{x}^I were extracted from wildlife scenes. The obtained images were then analyzed by a complex-cell filterbank $\mathcal{C}_{(x,y,\theta,f)}$ and the obtained responses were normalized. Then, the hidden redundancies were made explicit by FastICA.

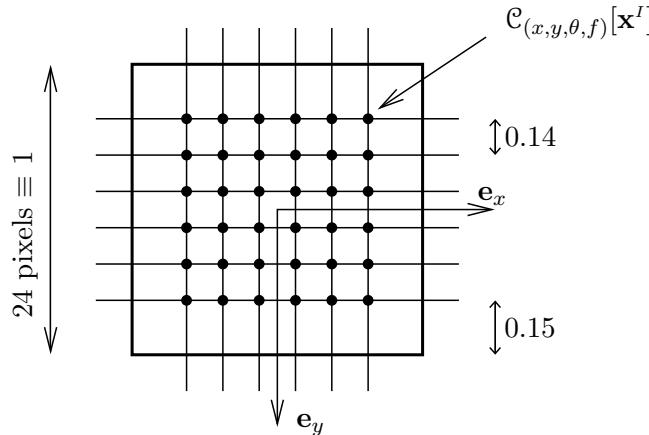


Figure 4.2: Filterbank which was used to extract image statistics. The size of the image was 24×24 pixels which defines one window unit. All remaining spatial quantities are given in window units. We placed complex-cells on each node of the grid. Each complex-cell response $\mathcal{C}_{(x,y,\theta,f)}[\mathbf{x}^I]$ was modeled with the energy model (2.23). The parameters x, y indicate the center position, θ the orientation and f the frequency tuning of the underlying Gabor filters as shown in Figure 2.5. Aspect ratio (2.20) was fixed to $l = 1.5$, frequency bandwidth (2.21) to $\Delta_f = 1.5$ oct. Peak spatial frequency was $f = 0.22$ cycles/pixel and $\theta = k \cdot \frac{\pi}{4}$, $k = 0, \dots, 3$. This led to Gabor filters with orientation bandwidth (2.22) $\Delta_\theta = 37^\circ$, length (2.18) $l_y = 0.26$ and width (2.18) $l_x = 0.17$.

$g_1(y) = \tanh(y)$	log of a sparse density.
$g_2(y) = y \exp[-\frac{y^2}{2}]$	log of a highly sparse density.
$g_3(y) = y^2$	Measures asymmetry of the data.
$g_4(y) = \exp\left[-\frac{y^2}{2}\right]$	Robust variant of g_3 .
$g_5(y) = \begin{cases} 3/2 \tanh(y) & \text{if } y > 0 \\ 1/2 \tanh(y) & \text{if } y < 0 \end{cases}$	Combination of g_3 and g_1 .

Table 4.1: Nonlinearities g which we have used in FastICA. g is the derivative of the function G that has been interpreted in Section 2.2 as the logarithm of a probability density and as a measure of non-Gaussianity. g_1 is the simplest choice. It assumes that the independent components s_i^H follow a sparse distribution, i.e. it implements a sparse prior. It can be seen as a robust variant of the log laplacian probability density. g_3 measures skewness of the data. g_5 combines skewness measurement and implementation of a sparse prior.

such that spatial and frequency overlap of the various filters was neither too big nor too small. It was a design problem and our choice is shown in Figure 4.2.

The nonlinearity g of the FastICA algorithm affects the learning of the mixing model in (4.4). In Table 4.1, it is noticed that different nonlinearities analyze different aspects in the data. That is, when we used the different nonlinearities in Table 4.1 to learn the mixture model, we maximized (2.11) either skewness or sparsity, or a combination of them.

While the Gabor filters and the nonlinearity g are used to measure statistical properties, dimension reduction can be used to generate good working-conditions for the learning process. Dimension reduction is linked to noise reduction and to low pass filtering. On the one hand, it may help to learn the mixing model, but on the other hand it is not any more possible to relate the learned model completely to maximization of non-Gaussianity or to the maximization of the logarithm of a probability density (2.11).

4.2.3 Interpretation of the ICA model

The ICA model in (4.4) is linear and generative. As discussed in Subsection 2.4.2, both s_i^H and x_i^C is thus interpreted as deviation of the firing rate of a neuron from its average. Assume the firing rate s^H of higher-order cell \mathcal{H} changes from zero to one. This would cause a change in the firing rate \mathbf{x}^C

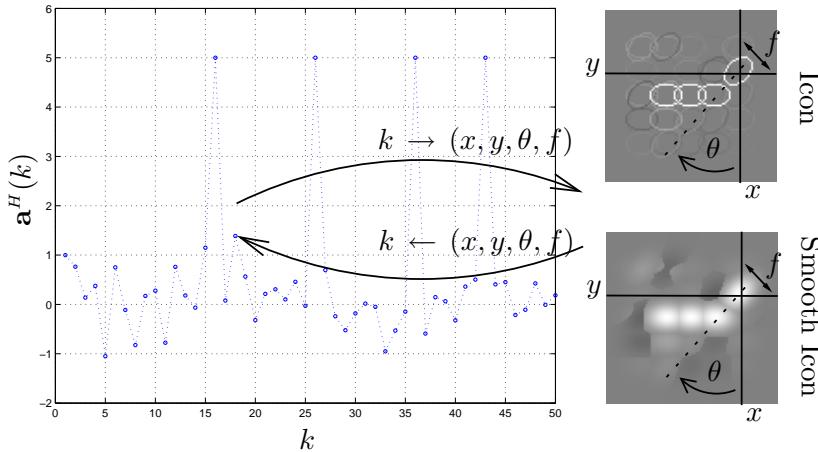


Figure 4.3: Higher-order cell \mathcal{H} is described by means of the generative weight vector \mathbf{a}^H . Visualization was achieved by plotting each component $\mathbf{a}^H(k)$ of the vector versus its index $k = k(x, y, \theta, f)$. Alternatively, the parameters (x, y, θ, f) were used to define a complex-cell icon centered at (x, y) , oriented to θ and of size proportional to $1/f$. Brightness of the icon was chosen as an indicator for the value of $\mathbf{a}^H(k)$, $k = k(x, y, \theta, f)$. Bright shading indicates a positive generative weight, dark shading a negative weight and gray is zero. The smooth icon is given by the “receptive field” of the complex-cell. In case of spatial overlap of the icons, the icon with the locally stronger weight was chosen to be depicted in the overlap region.

of all complex-cells. The amount of change would be given by the generative weight vector \mathbf{a}^H (4.4). This means that s^H codes for complex-cell activity as given by \mathbf{a}^H . We consider \mathcal{H} to represent, or code, that kind of activity.

If the k th component $\mathbf{a}^H(k)$ of the generative weight is positive, the effect of the change in firing rate s_i^H on $\mathbf{x}^C(k)$ is excitatory. If negative, the change in firing rate of \mathcal{H} has inhibitory influence.

In Subsection 4.2.1, we have denoted the output of a complex-cell at location (x, y) , tuned to orientation θ and frequency f by $\mathbf{x}^C(k)$, where k is the unique label for that cell. This means that the generative weight \mathbf{a}^H is a vector valued function of four parameters (x, y, θ, f) . For visualization of the generative weight, we chose to plot each component of the generative weight vector versus the index $k = k(x, y, \theta, f)$. Or alternatively, the parameters (x, y, θ, f) were used to define a complex-cell icon centered at (x, y) , oriented to θ and of size proportional to $1/f$. Brightness of the icon was chosen as an indicator for the value of $\mathbf{a}^H(k)$ with $k = k(x, y, \theta, f)$. See Figure 4.3.

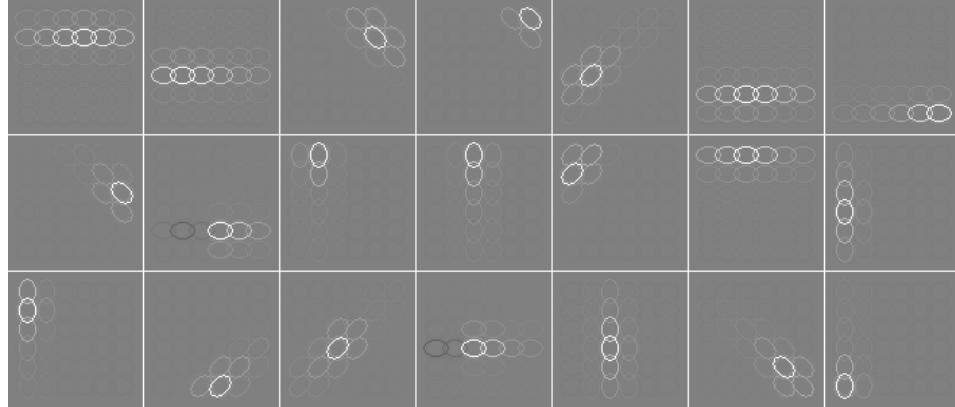


Figure 4.4: Random selection of learned generative weights \mathbf{a}_i^H . ICA nonlinearity g was the tanh nonlinearity g_1 . No normalization, i.e. $\alpha = 1$. Other FastICA nonlinearities led to similar generative weights as it is presented in Subsection 4.3.1.

4.3 Results

A linear representation of complex-cell outputs \mathbf{x}^H was learned by standard ICA to make structure in the complex-cell outputs explicit. Standard ICA yielded generative weights \mathbf{a}_i^H which are qualitatively similar to the weights learned by non-negative sparse coding in Section 3.4. That is, higher-order cells \mathcal{H}_i have emerged which code for simultaneous activity of collinear complex-cells, i.e. complex-cell contours.

Figure 4.4 shows a random selection of higher-order cells for the ICA nonlinearity g_1 and the normalization exponent $\alpha = 1$.

Organization of this section is as follows. In Subsection 4.3.1, we present how this finding depends on the choice of the FastICA nonlinearity g . In Subsection 4.3.2, we present the influence of normalization and dimension reduction.

4.3.1 Varying the learning criterion

The generative weights \mathbf{a}_i^H are qualitatively not different from those shown in Figure 4.4 if the learning process was driven by another FastICA nonlinearity g in Table 4.1 than g_1 . Length of the complex-cell contours, i.e. contours formed by simultaneously active complex-cells, has been measured based on methods in [13]. A Gaussian was fitted to match the values of the generative

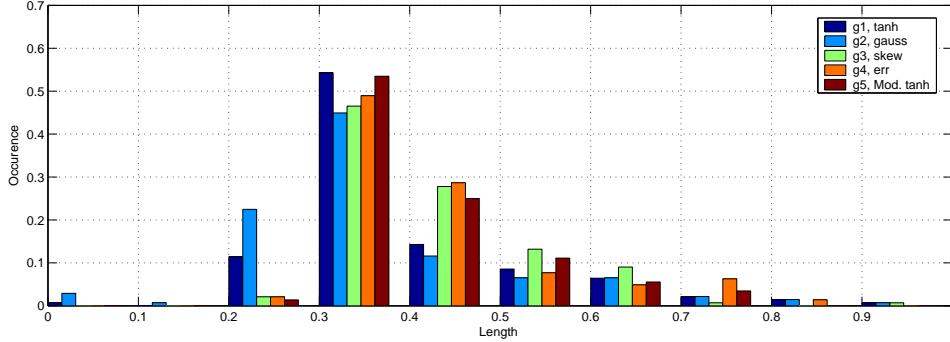


Figure 4.5: Distribution of pattern length for different FastICA nonlinearities g . Length was measured as given in (2.18). Deep blue: tanh nonlinearity g_1 (median of length: 0.38), light blue: gauss nonlinearity g_2 (median: 0.36), green: skew nonlinearity g_3 (median: 0.4) , orange: err nonlinearity g_4 (median: 0.40) and brown: modified tanh nonlinearity g_5 (median: 0.40). See Table 4.1 for a discussion of the FastICA nonlinearities.

weights. The effective length l_y of the fitted Gaussian was calculated as in (2.18). Figure 4.5 gives the length distribution for the tested FastICA nonlinearities. Nonlinearities which optimize skewness led to slightly longer complex-cell contours than nonlinearities which optimize sparseness only.

4.3.2 Varying preprocessing

Influence of dimension reduction on the generative weights is illustrated in Figure 4.6. The results are shown for dimension reduction from 144 dimensions to 80 (92% of original variance), to 60 (86% of original variance) and to 30 dimensions (66% of original variance). Length distribution of the learned patterns is given in Figure 4.7. Length of the complex-cell contours was found to increase with dimension reduction.

Influence of normalization with f_α or f_β is illustrated in Figure 4.8. We see that normalization with the values $\alpha = 0.5$ and $\beta = 0.5$ has led to end-stopped complex-cell contours.

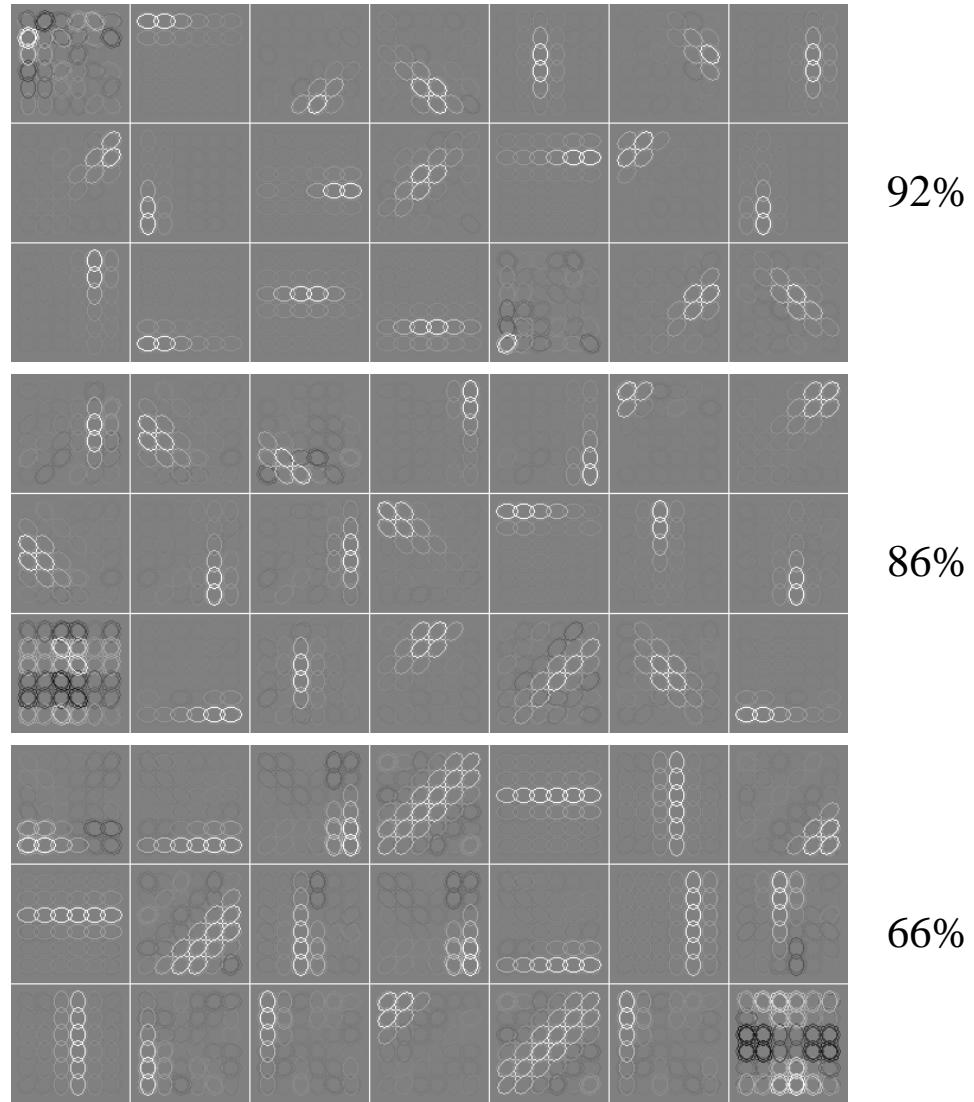


Figure 4.6: Random selection of learned generative weights \mathbf{a}_i^H for different amounts of dimension reduction. Percentage value on the right hand side indicates the amount of original variance which is retained.

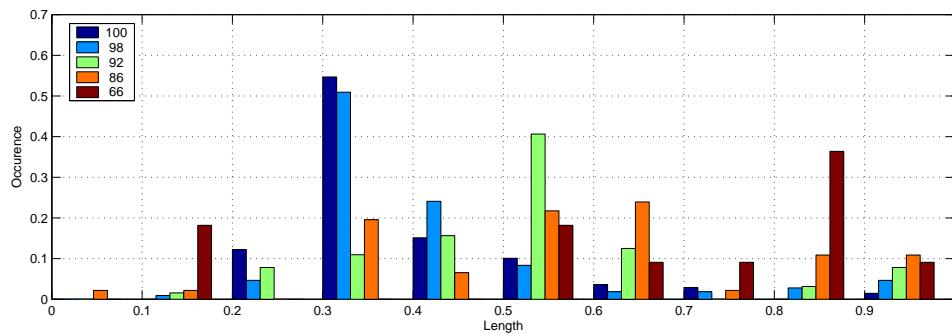


Figure 4.7: Distribution of complex-cell contour length for different degrees of dimension reduction. Length is measured as in (2.18). Deep blue: no dimension reduction, light blue: 98 % of original variance is retained, green: 92 % is retained, orange: 86% is retained, brown: 66% is retained.

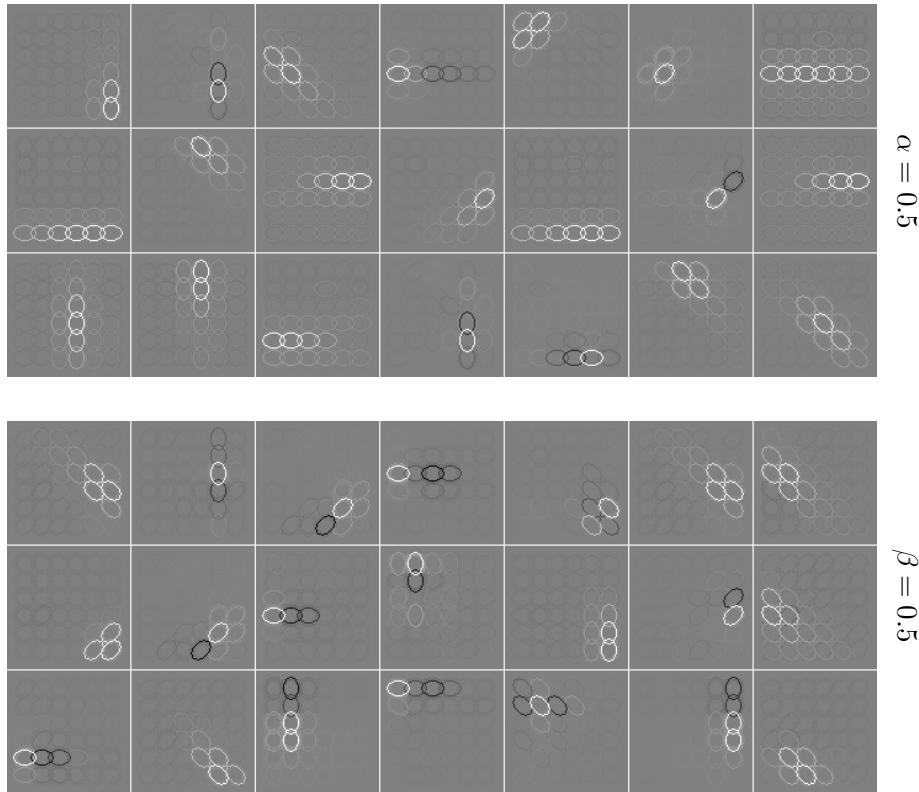


Figure 4.8: Random selection of learned generative weights \mathbf{a}_i^H for normalization with f_α and f_β .

4.4 Conclusion

We have addressed the research question of Subsection 3.5.1: We have investigated the necessity of the non-negativity constraint for contour cells to emerge. We conclude from Figure 4.4 that the representation of complex-cell outputs learned by standard ICA is qualitatively the same as the representation learned by non-negative sparse coding.

- We have shown in Figures 4.4 and 4.5 that both optimization of sparseness and optimization of skewness has led to contour cells as in [13].
- From Figure 4.5, we conclude that the learned representation is fairly robust with respect to the used FastICA nonlinearity.
- Complex-cell contours found by increased dimension reduction were shown in Figures 4.6 and 4.7 to be longer, but also not as well defined.
- We conclude from Figure 4.8 that normalization yields more sophisticated generative weights, including end-stopped higher-order cells.

Chapter 5

The Case of Multiple Scales

5.1 Introduction

In this chapter, we present investigations into the structure inherent in complex-cell outputs in the case that the cells are not all tuned to the same frequency. The motivation for this investigation has been formulated in Subsection 3.5.2. It included the evidence that according to Fourier theory, multiple frequencies are needed to describe real-world structures. Moreover, we stated that cells in V1 are continuously tuned in spatial frequency.

First, we explain how we have enhanced the experimental design of Chapter 4 to include multiple frequencies. Then, we present results which show that there are inter-scale dependencies, and we show how this structure is made explicit by ICA. Subsequently, we look at the significance of considering multiple frequencies in the experimental design. We show how the learned ICA model has incorporated complex-cell properties which are related to different scales.

5.2 Methods

The applied methods are based on the single-frequency methods of Chapter 4. The experimental design for that case has been presented in Section 4.2. We indicate here how it was extended to multiple frequencies.

At each node in Figure 4.2, we added complex-cells tuned to other frequencies than $f = 0.22$ cycles/pixel. Orientation tuning θ , frequency bandwidth Δ_f , aspect ratio l were not changed. The Gabor filters which underlie the complex-cells were designed such that their frequency peaks were loga-

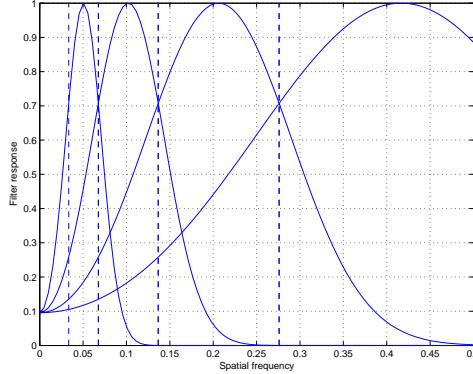


Figure 5.1: The experimental design as given in Section 4.2 was enhanced to include multiple frequencies. At every node in Figure 4.2 and every direction θ , we placed complex-cells tuned to four different frequencies. The underlying Gabor filters had logarithmically spaced frequency peaks and their frequency responses overlapped at the -3dB points. Peak spatial frequencies were chosen as follows: $f_1 = 0.05$ cycles/pixel, $f_2 = 0.1$ cycles/pixel, $f_3 = 0.21$ cycles/pixel and $f_4 = 0.42$ cycles/pixel.

rithmically spaced and that the frequency responses had overlaps at their -3dB points. Figure 5.1 shows the filter choice in the frequency domain.

Low frequency Gabor filters have a larger support than Gabors tuned to higher frequency. To balance numerically the filter responses, we used filters normalized to unit norm. By Parceval's theorem, this is equivalent to constraining the filters to capture the same amount of spatial energy.

We forwent normalization, i.e. $\alpha = 1$, and worked with the tanh FastICA nonlinearity g_1 without dimension reduction. See Subsection 4.2.1.

5.3 Results: Existence of dependencies

We show here that there are dependencies in complex-cell outputs across spatial frequency, or scale. Dependencies are visualized in Figure 5.2. The probability P in Figure 5.2 is the probability for a complex-cell to fire highly. It was determined by

$$P \left(\mathbf{x}^C(k) > \text{E}[\mathbf{x}^C(k)] + 4\sqrt{\text{var}[\mathbf{x}^C(k)]} \right). \quad (5.1)$$

The average value of P over all complex-cells was found as $P = 0.01$ with a standard deviation $\Delta_P = 0.001$. We calculated the conditional probability

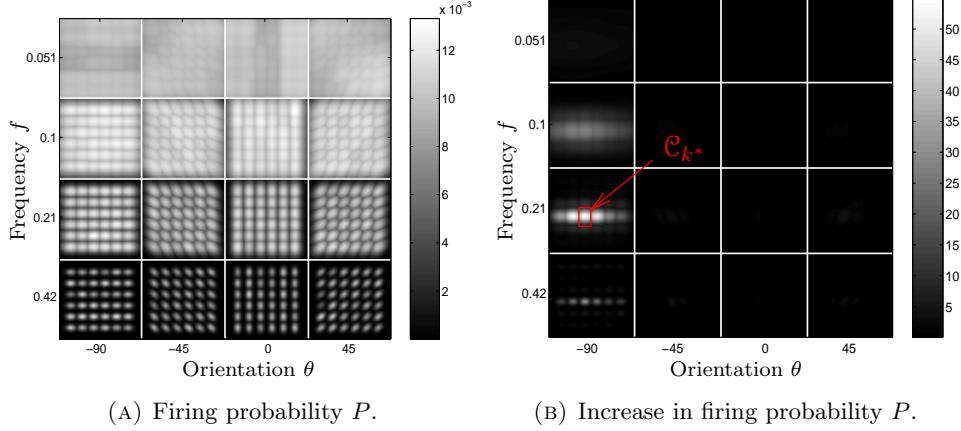


Figure 5.2: Visualization of dependencies in complex-cell outputs. A complex-cell is shown as a smooth icon and its shading indicates firing probability P , increase in P respectively. The firing-probability P for neighboring complex-cells of the same orientation increases for all scales if \mathcal{C}_{k^*} is firing. Maximal increase is around 50 times the normal probability to fire highly.

for each complex-cell to fire given that \mathcal{C}_{k^*} in Figure 5.2 is firing. This figure shows that the ratio between the conditional probability and P is high for neighboring complex-cells of the same orientation, across spatial scale. There are thus inter-scale dependencies.

Furthermore, Figure 5.3 shows that concurrent activity of two collinear, non-adjacent, complex-cells \mathcal{C}_1 and \mathcal{C}_2 has a facilitative effect on the latter ratio for equally oriented cells which are located in between \mathcal{C}_1 and \mathcal{C}_2 .

5.4 Results: Higher-order cell properties

Results in Section 5.3 indicate that there are dependencies between complex-cell outputs across the spatial scale. In this section, we show how we can find by ICA a representation of complex-cell outputs which make this hidden structure explicit. In other words, we show properties of the learned higher-order cells \mathcal{H}_i .

In Subsection 5.4.1, we present an overview of the learned generative weights, in Subsection 5.4.2, we highlight some of their properties. We visualize in Subsection 5.4.3 the generative weights by means of image synthesis.

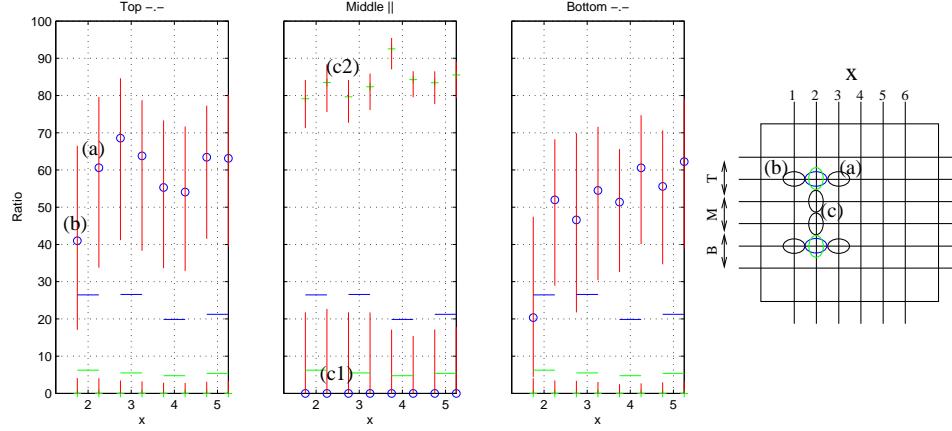


Figure 5.3: Concurrent activity of two collinear, non-adjacent, complex-cells increases substantially the probability P of firing for cells located in between. Values on the x -axis on the left hand side indicate the edge where the colored cells on the right hand side are centered. Assume the two blue-colored cells are firing at the same time. Increase in firing probability P of cell (a) on the right hand side is given by the blue circle labeled (a) on the left hand side. Similarly for cell (b). The blue circle labeled (c1) indicates the increase of P for cell (c) in case of concurrent activity of the blue-colored cells. The green circle (c2) indicates the increase if the two green-colored cells are firing together. Note that in that case the probability for cell (c) to fire is about 80 times the normal probability to fire. The ratio in Figure 5.2 where only one cell is assumed to fire is, in contrast, around 50.

5.4.1 Generative weight structure - overview

A random selection of generative weights \mathbf{a}_i^H is shown in Figure 5.4 in smooth icon form as introduced in Subsection 4.2.3. This representation suggests that generative weighting occurs in a coherent way across the different frequencies.

To investigate the coherence of weighting, let us define for any higher-order cell \mathcal{H}_i the hot-spot \mathbf{x}_i^* ,

$$\mathbf{x}_i^* = (x_i, y_i)^*, \quad (5.2)$$

as the center location (x, y) of complex-cells in Figure 4.2 where the average generative weighting by \mathcal{H}_i is highest.

$$\mathbf{x}_i^* = \arg \max_{\mathbf{x}} \sum_{\theta f} \mathbf{a}_i^H(k(\mathbf{x}, \theta, f)). \quad (5.3)$$

The preferred orientation θ_i^* of \mathcal{H}_i is then defined as

$$\theta_i^* = \arg \max_{\theta} \sum_f \mathbf{a}_i^H(k(\mathbf{x}_i^*, \theta, f)). \quad (5.4)$$

For 365 of 566 learned generative weights (65 %), complex-cells which are located at \mathbf{x}_i^* and tuned in orientation to θ^* were found to be the maximally weighted cells for *every* frequency f . This shows how generative weights pinpoint inter-scale dependencies. Furthermore, note that the locations \mathbf{x}_i^* appear as the bright spots in Figure 5.4, and that θ^* indicates overall orientation of the shown patterns.

5.4.2 Generative weight structure - by example

We detail here the properties of \mathcal{H}_{450} and \mathcal{H}_{380} which are shown in smooth icon representation in Figure 5.5.

For that purpose, we show the sub-vectors formed by the values $\mathbf{a}^H(k)$, $k = k(x, y, \theta_0, f_0)$ for fixed θ_0 and f_0 , separately in smooth icon representation. This representation reveals in Figure 5.6 that \mathcal{H}_{450} is coding for simultaneous across-scale activity of collinear complex-cells. Comparison of Figure 5.6 by Figure 5.2 shows how structure in complex-cell responses as indicated by the increase in firing probabilities is made explicit with ICA.

Analysis of the numerical values of the generative weights \mathbf{a}_{450}^H and \mathbf{a}_{380}^H further shows in Figure 5.7 that \mathcal{H}_{450} implements cross-orientation inhibition and \mathcal{H}_{380} inhibition across scales.

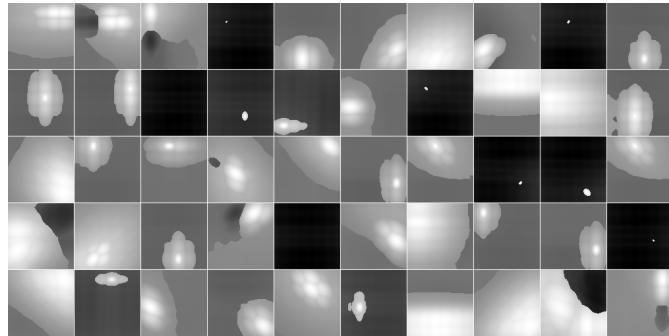


Figure 5.4: A random selection of generative weights \mathbf{a}_i^H in smooth icon representation (see Subsection 4.2.3). The location of brightest shading is \mathbf{x}_i^* and each pattern is oriented towards θ^* .

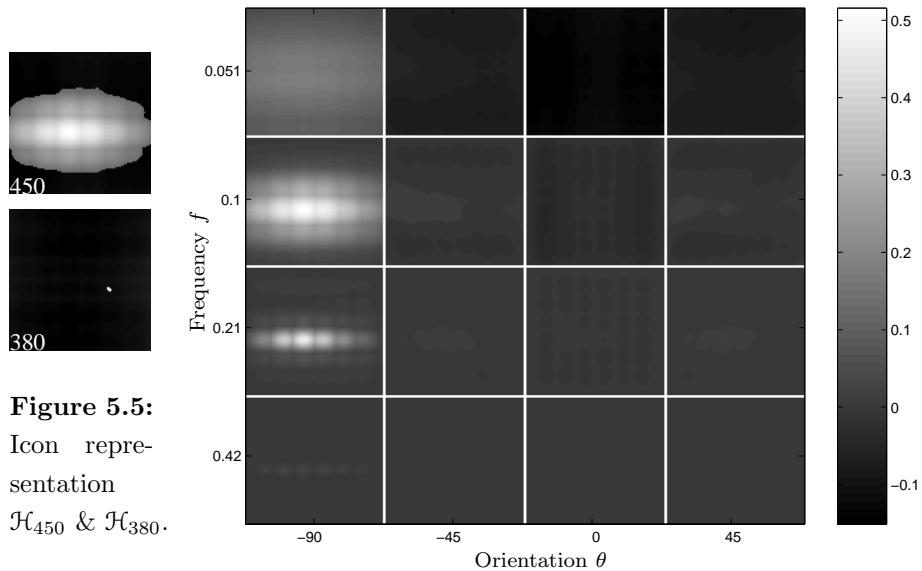


Figure 5.5:
Icon repre-
sentation
 \mathcal{H}_{450} & \mathcal{H}_{380} .

Figure 5.6: Icon representations of the sub-vectors formed by the values $\mathbf{a}_{450}^H(k)$, $k = k(x, y, \theta_0, f_0)$ for fixed frequency-orientation pairs (f_0, θ_0) .

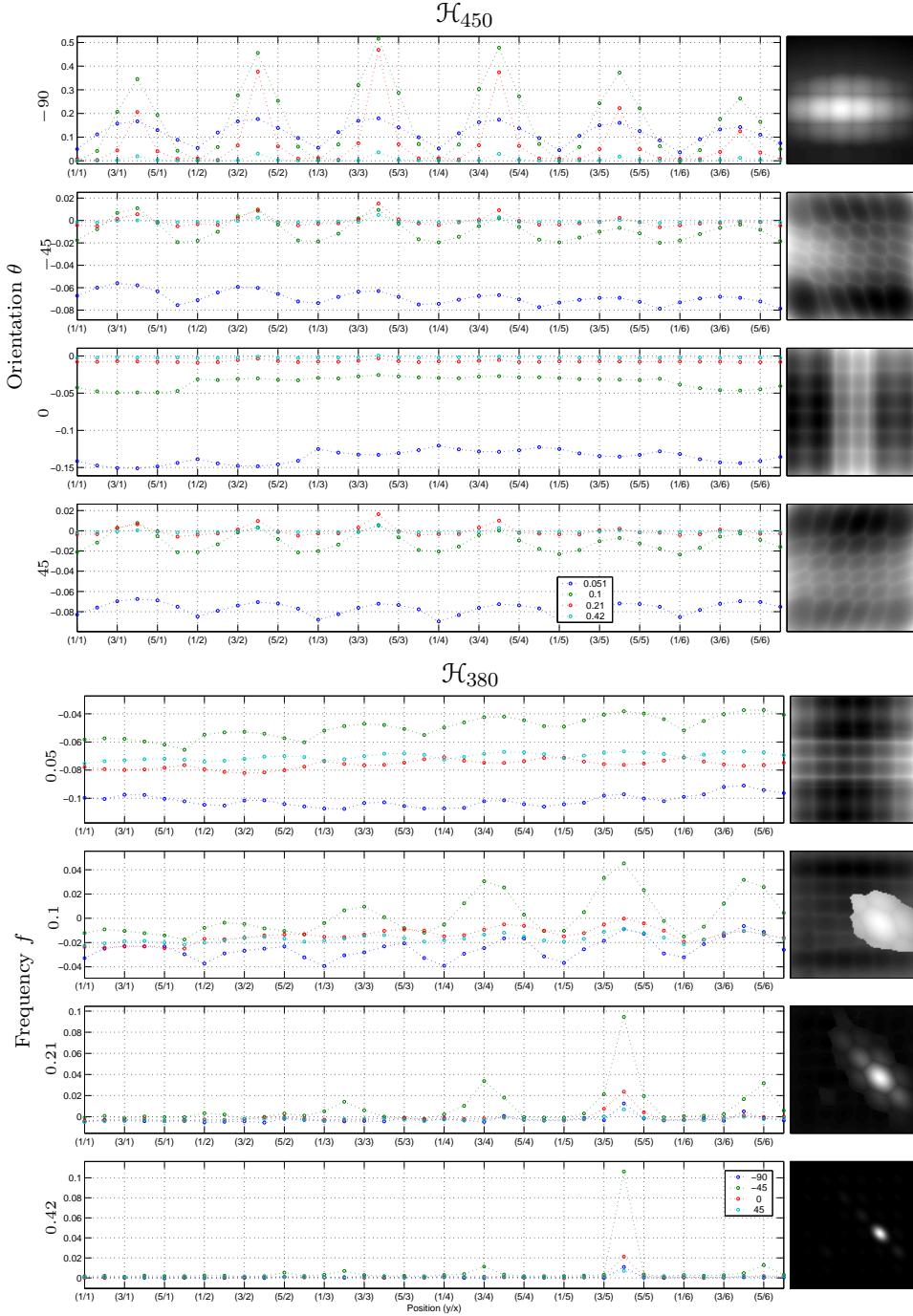


Figure 5.7: Generative behavior of sample higher-order cell \mathcal{H}_{450} (top) and \mathcal{H}_{380} (bottom). Cell \mathcal{H}_{450} shows cross-orientation inhibition. It inhibits complex-cells unless they are horizontally oriented. Cell \mathcal{H}_{380} shows inhibition across frequencies. Low frequency complex-cells are inhibited regardless of their location or orientation. Note the different scales. In each sub-plot, black stands for the lowest value of the data, bright for the highest.

5.4.3 Higher-order cells in the network

Figure 5.7 describes the higher-order cells \mathcal{H}_{450} and \mathcal{H}_{380} individually, i.e. isolated from other cells \mathcal{H}_i . We contrast them here with other cells to obtain a description of what they are coding for. We synthesize for that purpose an image from higher-order cell activities s_i^H as it is shown in Figure 5.8. Unlike in the previous section, we are thus not looking what the \mathcal{H}_i are coding for on the \mathcal{C} -layer, but on the image layer.

Complex-cell outputs were calculated based on the energy model as described in Section 4.2. The mapping from image to complex-cell outputs is not one-to-one. This means that the generation of the image is not uniquely defined given the activities s_i^H of the \mathcal{H}_i alone. A unique definition was achieved by constraining the symmetry of the images. Odd symmetric images were synthesized with

$$\mathbf{x}^C(k) = \sum_i \mathbf{a}_i^H(k) s_i^H + b, \quad (5.5)$$

where b is a bias term, and

$$\mathbf{x}^I = \sum_k \underbrace{\mathbf{w}_{o_k}^T \mathbf{x}^I}_{\sqrt{\mathbf{x}^C(k)}} \mathbf{w}_{o_k}. \quad (5.6)$$

The bias term b accounts for the fact that our linear model gives the deviation of the firing rate from the average (see Subsection 4.2.3). We have used in the latter equation vectorized odd Gabor filters \mathbf{w}_{o_k} as given in (2.17) instead of generative weights \mathbf{a}_k^S . This is an approximation. It can be justified by the similarity between the feed-forward weights \mathbf{w}^S and the generative weights \mathbf{a}^S of the simple-cell layer as shown in Figure 3.1.

Synthesis results are given in Figure 5.9. Images which were generated by random activity of the \mathcal{H}_i were taken as a baseline. Figure 5.9 shows the deviation of the generated images from the baseline for high activity of \mathcal{H}_{450} or \mathcal{H}_{380} . The latter cell codes for an edge-like image while the former cell codes for a notch.

5.5 Results: Significance of multiple scales

In Section 5.3 we have shown that there are dependencies in the outputs of complex-cells across spatial scale. In Section 5.4 we have shown that these

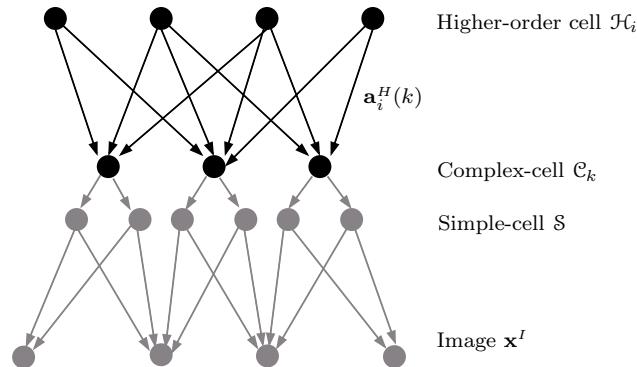


Figure 5.8: Hierarchical model used for image synthesis. The generative weights \mathbf{a}_i^H were learned as explained in Section 4.2. The grayed out layers were implemented by the energy model, see e.g. (2.23).



Figure 5.9: The effect of \mathcal{H}_{450} and \mathcal{H}_{380} reflects the structure of the learned generative weights in Figure 5.7. Activity of \mathcal{H}_{450} leads to an edge-like image, activity of \mathcal{H}_{380} to a notch.

hidden redundancies were pinpointed by ICA. Learning a linear representation of complex-cell outputs from natural images has led to higher-order cells which code for concurrent across-scale activity of complex-cells. In terms of images, we have seen that they can code for edges or notches. In this section, we look at the significance of having taken multiple scales into account.

First, we show how the presence of multiple frequencies leads to the generation of a real-world edge. We performed a local image synthesis from the generative weights of \mathcal{H}_{450} at its hot spot \mathbf{x}_{450}^* . Figure 5.10 shows a slice orthogonal to the preferred orientation θ_{450}^* . The intensity of the synthesized image shows no side-lobes whereas the underlying Gabor filters do. This means the synthesized curve flattens gradually out without going negative. This shape corresponds more to the notion of an edge than a Gabor filter does.

Next, we consider a higher-order cell as an edge detector. Edge detectors have briefly been discussed in Subsection 2.5.2 and Subsection 2.5.3. We have stated advantages of a quadratic filter over a linear filter and pointed out that for single-scale filters good localization performance and good noise suppression were conflicting. Broad filters have a good signal to noise ratio (SNR), but they are not good at localizing edges. Vice versa for narrow filters. It was mentioned that multi-scale filters might be able to resolve that conflict. The feed-forward weights \mathbf{w}_i^H which were learned by ICA implement actually multi-scale quadratic filters (Subsection 2.5.3). Response s_i^H of higher-order cell \mathcal{H}_i is the output of the filter. Figure 5.11 shows that these filters combine localization performance of high-frequency Gabors with noise suppression performance of broad Gabors.

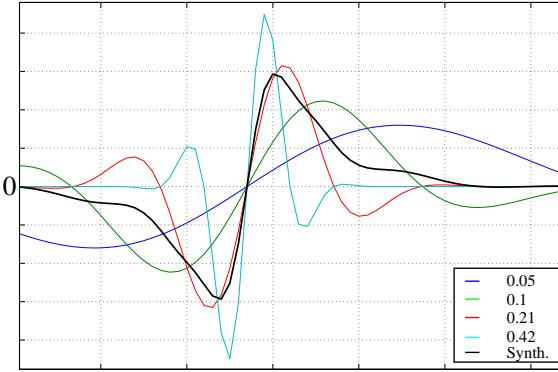


Figure 5.10: Local image synthesis from the four generative weights $\mathbf{a}_{425}^H(k^*)$, $k^* = k(x^*, y^*, \theta^*, f)$. Slice orthogonal to θ^* . The synthesized curve corresponds more to the notion of an edge than a Gabor filter does. See Figure 5.9 for an image synthesis which takes all generative weights \mathbf{a}_{450}^H into account.

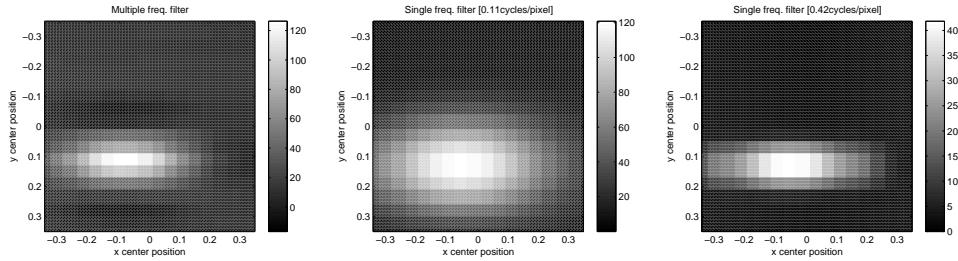


Figure 5.11: SNR for a narrow bar stimulus. Center position of the bar is given by the coordinates on the x-axis and y-axis. Left: Multi-scale quadratic filter implicitly defined by \mathcal{H}_{450} as learned by ICA from natural images. Middle: Quadratic filter given by collinearly arranged complex-cells tuned to $f = 0.1$ cycles/pixel. Right: Idem for complex-cells tuned to $f = 0.42$ cycles/pixel. The broad single-frequency filter is good at noise suppression, not good, however at localization. Vice versa for the high-frequency filter. The learned multi-scale filter combines the SNR of the broad single-frequency filter and with localization performance of the high-frequency filter. Note different scales.

5.6 Conclusion

We have addressed the research question of Subsection 3.5.2: We have investigated structure which is contained in the outputs of complex-cells which are not all tuned to the same spatial frequency.

- We have shown in Section 5.3 that there are dependencies between complex-cell outputs across the spatial scale. Dependencies are similar on every scale as predicted by the scale-invariance hypothesis stated in Subsection 2.3.2.
- Hidden redundancy reduction by ICA leads to a linear representation of complex-cell outputs which makes this structure explicit (Section 5.4). We have shown in Figure 5.6 how a higher-order cell \mathcal{H} codes for simultaneous across-scale activity of collinearly arranged complex-cells.
- We have interpreted in Subsection 5.4.3 higher-order cell activity in terms of images. We have seen in Figure 5.9 and Figure 5.10 how coherent generative weighting across the frequencies leads to real-world images, such as an edge or a notch.
- Figure 5.7 shows that cross-orientation inhibition and cross-frequency inhibition has been learned from natural images. Whereas inhibition across frequencies has been identified in Subsection 2.4.2 as a possible mechanism which leads to end-stopped cells.
- By making structure in complex-cell outputs explicit, we have implicitly designed a quadratic multi-scale filter. This filter combines properties which are conflicting for single-scale filters, e.g. noise suppression and localization performance (Figure 5.11).

Chapter 6

Validity of the Experimental Design

6.1 Introduction

In Chapter 4 and Chapter 5, we have shown how hidden redundancy reduction by ICA made well-defined higher-order cells emerge. In this chapter, we present investigations into the validity of the involved experimental design. In Subsection 3.5.3, motivation was given to focus for that purpose on the influence which a filterbank has on the kind of feature which is learned by ICA. We analyze here to which extent ICA results do reflect properties of natural images and to which extent properties of the filterbank.

6.2 Methods

The adopted method consisted in contrasting results obtained by the filterbank used in the previous chapters with results obtained by a second filterbank. We have briefly described in Section 3.3 how complex-cell filters can be learned by ISA from natural images. We chose to use these filters for the latter filterbank.

6.2.1 Investigation by correlation analysis

First, we investigated the influence of the filterbank by means of correlation analysis. Denote by \mathbf{w}^C a ISA filter as in Section 3.3 and by \mathbf{w}_{G_i} a vectorized Gabor filter as defined in (2.17). For both ISA filters and Gabor filters,

we calculated the correlation coefficient r for collinearly arranged filters in function of distance and phase.

Gabor filters

Gabor filters \mathbf{w}_{G_i} were arranged in a collinear manner as shown in Figure 6.1. Correlation coefficient r between the response s_0 of the even Gabor filter \mathbf{w}_{e_0} ,

$$s_0 = \mathbf{w}_{e_0}^T \mathbf{x}^I, \quad (6.1)$$

and the response s_i of \mathbf{w}_{G_i} ,

$$s_i = \mathbf{w}_{G_i}^T \mathbf{x}^I, \quad (6.2)$$

was calculated by

$$r = \frac{\text{cov}[h(s_0), h(s_i)]}{\sqrt{\text{var}[h(s_0)] \text{var}[h(s_i)]}}. \quad (6.3)$$

We determined linear correlation, $h(y) = y$, correlation of squares, $h(y) = y^2$, and correlation between square root rectified outputs, $h(y) = \sqrt{|y|}$.

ISA filters

The used ISA filters were learned as explained in previous work which has been presented in Section 3.3. We had to determine key parameters of the ISA filter \mathbf{w}_i^C to be able to measure correlation r between their outputs in the same way as for Gabor filters.

- Length l_{x_i} (2.18), width l_{y_i} (2.19) and center location (x_i, y_i) (Figure 2.5) of the filters were calculated by fitting a Gaussian to the envelope of \mathbf{w}_i^C in the least squares sense. The envelope was calculated by taking the 2-dimensional Hilbert transform of \mathbf{w}_i^C , followed by the square operation on both \mathbf{w}_i^C and its Hilbert Transform, adding them together and finally taking the square root [10].
- Peak spatial frequency f_i , orientation θ_i and phase ϕ_i , see (2.17), were determined by stimulating the filter \mathbf{w}_i^C with different sinusoidal inputs and retaining the parameters which yielded the maximum response.

For a fixed ISA filter \mathbf{w}_i^C , approximately collinear filters \mathbf{w}_j^C which had the same frequency tuning were determined. Filters were considered to be collinear to \mathbf{w}_i^C if they lay in a cone with apex (x_i, y_i) and apex angle

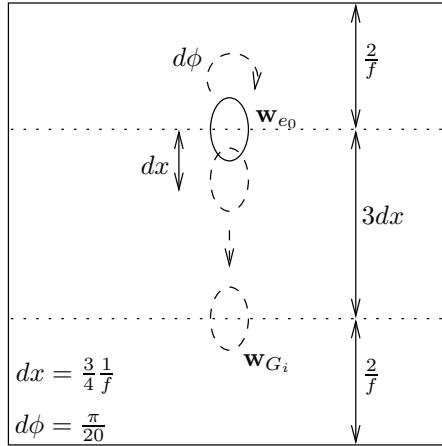


Figure 6.1: Determination of the correlation coefficient r for Gabor filters which were arranged in a collinear manner. All filters involved had orientation tuning (2.17) $\theta = 0$ rad., frequency tuning (2.17) $f = 0.21$ cycles/pixel, frequency bandwidth (2.21) $\Delta_f = 1.5$ oct. and aspect ratio (2.20) $l = 1.5$. The even Gabor filter \mathbf{w}_{e0} was located at a distance of two wavelengths from the image border. Gabor filters \mathbf{w}_{Gi} were shifted in phase by $d\phi = k_1 \frac{\pi}{20}$, $k_1 = 0, \dots, 19$ and placed at the same location, and at places $dx = k_2 \frac{3}{4} 1/f$ underneath, $k_2 = 1, \dots, 3$. Shift distance dx was $dx = 0.14$.

18° and if their orientation θ_j deviated no more than 9° from θ_i . Filters were considered to have the same frequency tuning, if their peak frequency deviated no more than 0.15 cycles/pixel from f_i .

Distance $d(i, j)$ between these filters was then calculated by

$$d(i, j) = \frac{\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}}{\sqrt{l_{y_i} l_{y_j}}}, \quad (6.4)$$

and correlation r was determined in function of distance d and difference in phase.

6.2.2 Investigation by ICA

Next, we investigated the influence of the filterbank by ICA. For that purpose, we used the single-frequency experimental design of Section 4.2. We did, however, not perform ICA on complex-cell responses $\mathcal{C}[\mathbf{x}^I]$ but on the squared responses of the underlying simple-cells. That is, we omitted the pooling stage in (2.23). Denote by \mathbf{x}^E the vector which collects all the

squared responses x_i^E ,

$$x_i^E = (\mathbf{w}_{e_i}^T \mathbf{x}^I)^2, \quad (6.5)$$

and by \mathbf{x}^O the vector which collects all

$$x_i^O = (\mathbf{w}_{o_i}^T \mathbf{x}^I)^2, \quad (6.6)$$

where \mathbf{w}_{e_i} is an even Gabor filter and \mathbf{w}_{o_i} an odd Gabor filter. We estimated then the ICA model

$$\begin{bmatrix} \mathbf{x}^O \\ \mathbf{x}^E \end{bmatrix} = \sum_i \mathbf{a}_i^H s_i^H \quad (6.7)$$

$$= \sum_i \begin{bmatrix} \mathbf{a}_i^{Ho} \\ \mathbf{a}_i^{He} \end{bmatrix} s_i^H \quad (6.8)$$

from natural images.

6.3 Results: Investigation by correlation analysis

Figure 6.2 shows correlation results for collinearly arranged Gabor filters as given in Figure 6.1. We see that outputs of Gabor filters are linearly correlated. For (square) rectification, we find that the correlation coefficient r is higher for an even Gabor filter which is not centered at the same location as \mathbf{w}_{e_0} than for an odd Gabor filter which is centered at the same location. That is, correlation results are strongly phase dependent. Note that similar results were found for Gabor filters tuned to other frequencies than $f = 0.21$, as predicted by the scale-invariance hypothesis given in Subsection 2.3.2.

Parameterization results for ISA filters are shown in Figure 6.3 and in Figure 6.4. The results are similar to those in [19] and in [10]. Figure 6.5 shows that ISA filters led to different correlation results than Gabor filters. First, ISA filters have linearly uncorrelated outputs. Secondly, the (square) rectified outputs of nearby filters are, regardless of their phase, more correlated than the (square) rectified outputs of equal-phase filters centered at different locations.

Since ISA filters in a common complex-cell subspace tend to differ in phase, but not in localization, orientation or frequency (Figure 6.3 or Section 3.3), the results shown in the left Sub-figure in Figure 6.5 are measurements of intra-subspace correlation. The correlation coefficient r is around

$r = 0.3$ for rectified outputs. We calculated the probability P for a ISA filter to have rectified outputs which are more correlated to outputs of filters exterior to the subspace than to those of filters inside the subspace. The probability P is $P = 0.03$, i.e. very low. In Section 3.3, we have introduced the notion of correlation in energy. We have defined two random variables to be correlated in energy if their (square) rectified outputs are correlated, not however their outputs themselves. With that definition, we deduce from the low value of probability P that correlation in energy is highest if the involved filters are in different phases, but of similar orientation, location and frequency.

In stark contrast, it is not possible to deduce this evidence from the measurement by Gabor filters, because in that case, correlation of squares is partly caused by linear correlation and it is not only due to correlation in energy, i.e. pure second-order correlation. Linear correlation has contaminated the measurement of correlation in energy. Indeed, the influence of linear correlation was seen to be such that correlation of squares becomes strongly phase dependent whereas correlation in energy was found to be not a sensitive function of phase.

6.4 Results: Investigation by ICA

In this section, we illustrate how properties of the filterbank affect ICA results. For that purpose, we will refer to results of the previous section. We performed ICA on squared responses x_i^E and x_i^O of even and odd Gabor filters. Given that filters in different phases dominate correlation in energy, it could have been expected that pooling of the squared responses x_i^E and x_i^O would be learned from natural images. However, this did not happen and the generative weights $\mathbf{a}_i^H = [\mathbf{a}_i^{Ho}; \mathbf{a}_i^{He}]^T$ were such that it was possible to separate them into two subsets. For a given \mathbf{a}_i^H ,

$$\mathbf{a}_i^H = \begin{bmatrix} \mathbf{a}_i^{Ho} \\ \mathbf{a}_i^{He} \end{bmatrix}, \quad (6.9)$$

an $\mathbf{a}_{i^*}^H$ was found to exist that verified

$$\mathbf{a}_{i^*}^H \approx \begin{bmatrix} \mathbf{a}_i^{He} \\ \mathbf{a}_i^{Ho} \end{bmatrix}. \quad (6.10)$$

That is, the two subsets coded for features of equal form, but of switched symmetries. An example is given in Figure 6.6. The generative weights \mathbf{a}_i^H

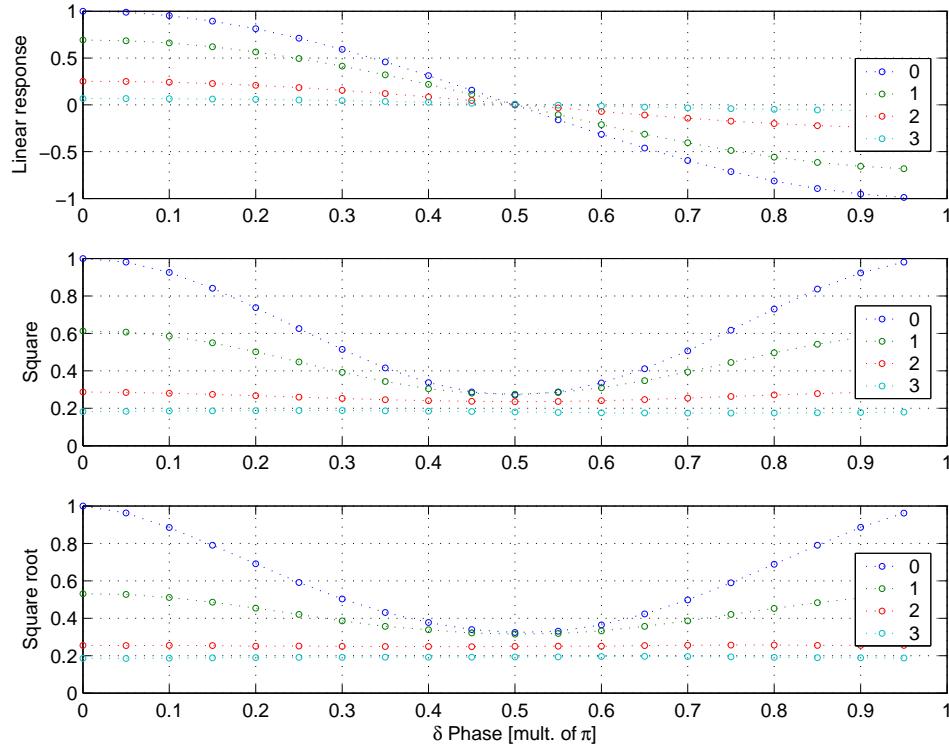


Figure 6.2: Correlation results for Gabor filters. Blue: no spatial shift, Green: vertical shift by dx , Red: vertical shift by $2 dx$, cyan: vertical shift by $3 dx$. Linear filter responses are correlated. Square and square root rectification yield similar results. Correlation between even and odd rectified responses are around $r = 0.3$ regardless of the vertical spatial shift. This is less than r for equal-phase filters spaced by dx and about the same as r for equal-phase filters spaced by $2 dx$.

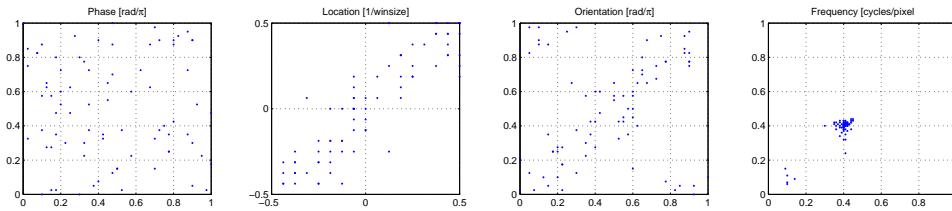


Figure 6.3: Local parameter correlation of ISA filters in a common complex-cell subspace. This result is a replication of the one shown in Figure 3.2, due to [19].

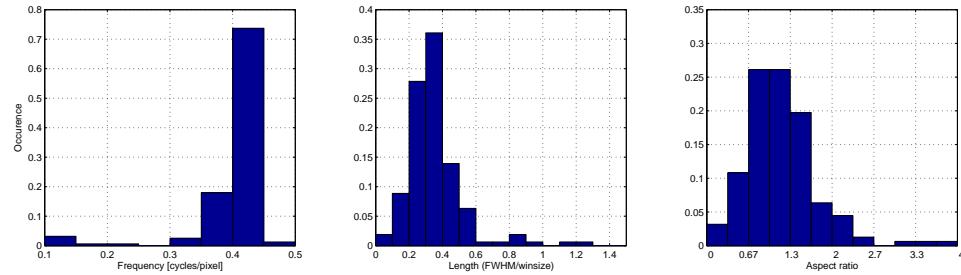


Figure 6.4: Spatial extension measurement of ISA filters. The results are similar to those in [10]. Median of the length (l) is $l = 0.32$, the 25% and 75% percentiles are $l = 0.26$ and $l = 0.41$. The length l of filters used in Chapter 4 is $l = 0.26$ and a complex-cell contour formed by two collinear complex-cells is $l = 0.39$ long.

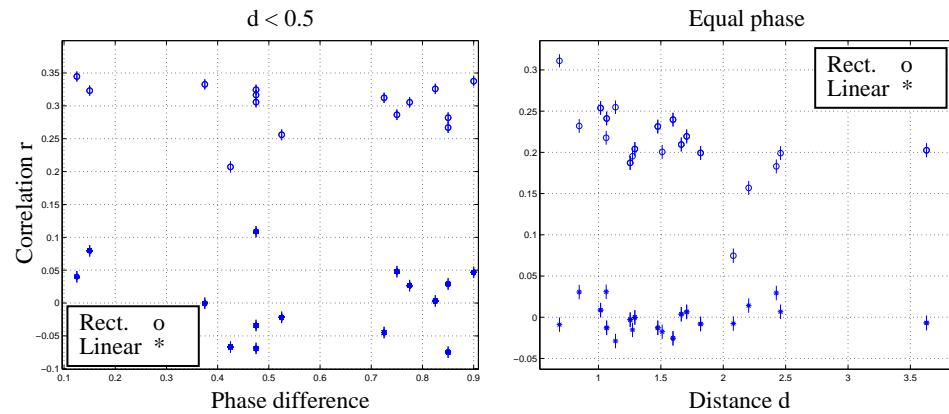


Figure 6.5: Correlation results for ISA filters. Left: Nearby filters. r is around $r = 0.3$ Right: Spatially spaced filters of approximately equal phase. r is around $r = 0.2$. The squares of nearby filters are slightly, but clearly, more correlated than the squares of further-away filters, regardless of their phase. Note the different scale.

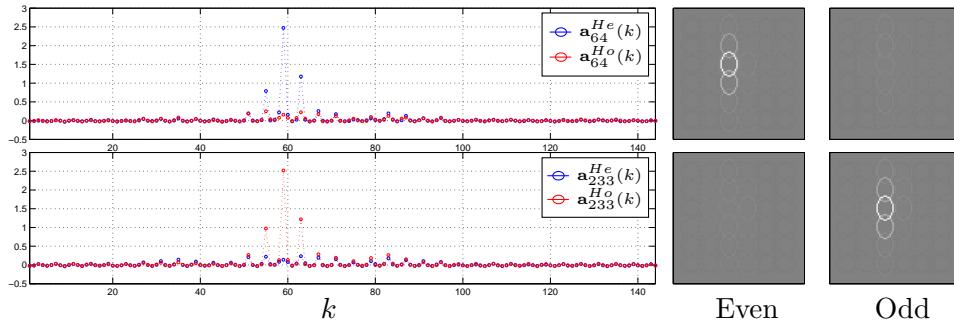


Figure 6.6: The two generative weights code for a feature of the same form. One however for a feature of even symmetry, the other for a feature of odd symmetry. ICA made thus symmetry in natural images explicit, not however dependencies between squared outputs of even and odd Gabor filters.

were such that they coded the simultaneous activity of collinearly arranged Gabors like in Chapter 4. However, the contours were of specific symmetry which reflects the correlation results in Figure 6.2.

Linear correlation has masked pure higher-order correlation so that dependencies due to the latter were not made explicit with ICA. But dependencies due to linear correlation dominate the learned representation. See Figure 6.7 for an illustration.

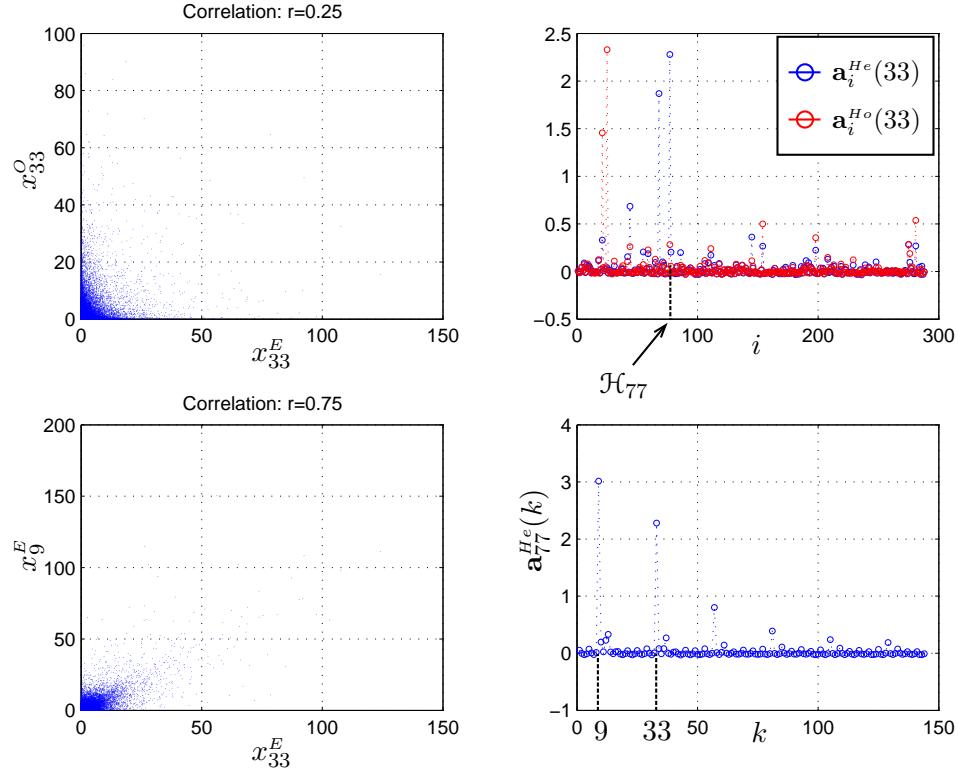


Figure 6.7: Influence of linear correlation on the learning of generative weights. Upper left: Scatter plots of the squared responses of $\mathbf{w}_{e_{33}}$ and $\mathbf{w}_{o_{33}}$, an even Gabor and an odd Gabor respectively. Correlation coefficient $r = 0.25$. Upper right: There is no higher-order cell \mathcal{H}_i which weights the even and odd Gabor strongly at the same time. Lower right: Generative weighting of \mathcal{H}_{77} , the higher-order cell which weights the even Gabor $\mathbf{w}_{e_{33}}$ strongest. Lower left: Scatter plots of the squared responses of $\mathbf{w}_{e_{33}}$ and $\mathbf{w}_{e_{19}}$. The latter Gabor is strongly weighted by \mathcal{H}_{77} and it is collinear to $\mathbf{w}_{e_{33}}$. Correlation coefficient $r = 0.75$. The four sub-figures show that dependencies due to pure correlation of squares were not made explicit by ICA, but linear correlation between collinear Gabors.

6.5 Conclusion

We have addressed the research question of Subsection 3.5.3: We investigated the validity of the experimental design by analyzing the influence of the filterbank on the estimates of the linear ICA model. We emphasize hereby that it is important to consider properties of the filterbank.

- We have shown in Section 6.3 how linear correlation can mask pure higher-order correlation. This can constitute an impediment to the task of testing hidden redundancy reduction on areas beyond V1 since these areas might just analyze higher-order dependencies.
- By example, we have shown in Section 6.4 that it is important to consider properties of the filterbank. Our intention was to learn pooling by ICA but we learned two sets of contour-cells, an even and an odd one (Figure 6.6). That we have not learned pooling, however, is not due to the wrong computational principle, i.e. independence maximization under a sparse prior (see Section 3.3), but it is due to the fact that the used filterbank was inappropriate for that task.

Chapter 7

Conclusions

In this thesis, we have analyzed hidden structure inherent in the response of quadrature phase filters to natural images. Our approach consisted in representing the responses in such a way that the structure becomes explicit. We used the concept of hidden redundancy reduction to find such a representation.

Hidden redundancy reduction is hypothesized to be a strategy employed by the visual cortex to discover structure in natural images. Since quadrature phase filters are a model for complex-cells, which are located in the primary visual cortex, research presented in this thesis can be considered to test the hidden redundancy reduction hypothesis on areas beyond the primary visual cortex.

7.1 Conclusions

We draw from our research the following conclusions.

1. Motivated by the research question formulated in Subsection 3.5.1, we have shown in Section 4.3 that, like non-negative sparse coding, standard ICA leads to a representation of complex-cell outputs which makes joint activity of collinearly arranged complex-cells explicit.
2. Motivated by the same question and by the question of Subsection 3.5.2, we have shown in Subsection 4.3.2 and in Subsection 5.4.2 how hidden redundancy reduction by standard ICA can account for end-stopping as described in Subsection 2.4.2. Moreover, we have presented in Subsection 5.4.2 that it can account for cross-orientation inhibition too.

3. Based on the research question formulated in Subsection 3.5.2, we have shown in Subsection 5.4.1 how standard ICA pinpoints across-scale dependencies in complex-cell responses to natural images. In Subsection 5.4.3 and in Section 5.5, we showed how these dependencies lead to a representation of complex-cell outputs which makes higher-order features like real-world edges or notches manifest.
4. Motivated by the question of Subsection 3.5.3, we have shown in Chapter 6 that the exact design of the quadrature phase filters used to test hidden redundancy reduction is of importance.

7.2 Discussion

We discuss here to which area in the visual cortex our results could be related to. We explain that it depends on the used quadrature phase filters whether hidden redundancy reduction was tested on areas beyond V1 or not.

Conclusions 1 and 2 In Chapter 4, we have used quadrature phase filters tuned to one frequency only. We have shown in Section 6.3 how in that case, higher-order dependencies between the filter outputs are masked by linear correlation. It is likely that this is a reason for the robust emergence of complex-cell contours as reported in Section 4.3. Given this influence of linear correlation, the filters might be more suitable to test hidden redundancy reduction on the primary visual cortex than beyond. The results in Chapter 4 could be seen as evidence for hidden redundancy reduction to account for length distribution and end-stopping behavior of cells in the primary visual cortex. That is, a complex-cell contour might be a merely *long* complex-cell.

Conclusions 2 and 3 In Chapter 5, the quadrature phase filters were of various scales. We relate results in Section 5.4 and Section 5.5 to areas beyond V1 because they involve an interplay of multiple frequencies whereas cells in the primary visual cortex are sharply tuned in frequency.

Bibliography

- [1] J.J. Atick. Could information theory provide an ecological theory of sensory processing. *Network*, 3:213-251, 1992.
- [2] Y.K. Aw, R. Owens, and J. Ross. Analysis of local energy and phase congruency models in visual feature detection. *J. Austral. Math. Soc. Ser. B* 40, 97-122, 1998.
- [3] B. Barlow. Redundancy reduction revisited. *Network: Computation in neural systems*, 12:241-253, 2001.
- [4] A.J. Bell and T.J. Sejnowski. The ‘independent components’ of natural scenes are edge filters. *Vision Research*, 37(23):3327–3338, 1997.
- [5] J.F. Canny. Finding edges and lines in images. Technical Report AI-TR-720, MIT Artificial Intelligence Laboratory, 1983.
- [6] M. Carandini, D.J. Heeger, and J.A. Movshon. *Cerebral Cortex, vol XII. Cortical Models*, chapter Linearity and Gain Control in V1 Simple Cells. Plenum Press, 1996.
- [7] P. Comon. Independent component analysis, a new concept? *Signal Processing*, 36: 287-314, 1994.
- [8] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc., 1991.
- [9] F. Gabbiani. Models of V1 simple and complex cells. Lecture Notes, Baylor College of Medicine 2003.
- [10] J.H. van Hateren and A. van der Schaaf. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc.R.Soc.Lond.B.*, 265:359-366, 1998.

- [11] P. Hoyer. Non-negative sparse coding. In *Proceedings IEEE Workshop on Neural Networks for Signal Processing*, pages 557–568, 2002.
- [12] P. Hoyer. *Probabilistic Models of Early Vision*. PhD thesis, Helsinki University of Technology, Neural Networks Research Centre, 2002.
- [13] P. Hoyer and A. Hyvärinen. A multi-layer sparse coding network learns contour coding from natural images. *Vision research*, 42(12):1593-1605 (2002), 2002.
- [14] D. Hubel. *Eye, Brain, and Vision*. Scientific American Library Series. New York: WH Freeman, 1988.
- [15] J. Hurri and A. Hyvärinen. Simple-cell-like receptive fields maximize temporal coherence in natural video. *Neural Computation*, 15(3): 663-691, 2003.
- [16] A. Hyvärinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3):626-634, 1999.
- [17] A. Hyvärinen. Independent component analysis: Algorithms and applications. *Neural Networks*, 13(4-5):411-430, 2000.
- [18] A. Hyvärinen and P. Hoyer. Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces. *Neural computation*, 12:1705-1720, 2000.
- [19] A. Hyvärinen and P. Hoyer. A two-layer sparse coding model learns simple and complex cell receptive fields and topography from natural images. *Vision Research*, 41: 2413-2423, 2001.
- [20] A. Hyvärinen, P. Hoyer, and M. Inki. Topographic independent component analysis. *Neural Computation*, 13(7):1527-1558 (2001), 2001.
- [21] A. Hyvärinen, J. Karhunen, and J.E. Oja. *Independent Component Analysis*. Wiley-Interscience, 2001.
- [22] J.P. Jones and L.A. Palmer. An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58(6):1233–1258, 1987.

- [23] C. Kayser, W. Einhäuser, O. Dümmer, P. König, and K. Körding. Extracting slow subspaces from natural videos leads to complex cells. In *Proceedings of the International Conference on Artificial Neural Networks (ICANN)*, pages 1075–1080, 2001.
- [24] D. Kersten and A. Yuille. Bayesian models of object perception. *Current Opinion in Neurobiology*, 13: 150-158, 2003.
- [25] P. Kovesi. Edges are not just steps. *ACCV2002, The 5th Asian Conference on Computer Vision*, 2002.
- [26] R. Mehrotra, K.R. Namuduri, and N. Ranganathan. Gabor filter-based edge detection. *Pattern Recognition*, Vol. 25, No. 12:1479-1494, 1992.
- [27] B.A. Olshausen. Principles of image representation in visual cortex. In L.M. Chalupa and J.S. Werner, editors, *The Visual Neurosciences*, pages 1603–15. MIT Press, 2002.
- [28] B.A. Olshausen and D.J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381: 607-609, 1996.
- [29] B.A. Olshausen and D.J. Field. Natural image statistics and efficient coding. *Network: Computation in neural systems*, 7:333-339, 1996.
- [30] B.A. Olshausen and D.J. Field. Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research*, 37:3311-3325, 1997.
- [31] S.E. Palmer. *Vision Science - Photons to Phenomenology*, chapter Processing Image Structure. MIT press, 1999.
- [32] P. Perona and J. Malik. Detecting and localizing edges composed of steps, peaks and roofs. In *Proceedings of the Third International Conference of Computer Vision*, pages 52–57, 1990.
- [33] E.P. Simoncelli and B.A. Olshausen. Natural image statistics and neural representation. *Annu.Rev.Neuroscience*, 24:1193:216, 2001.
- [34] A. Srivastava, A.B. Lee, E.P. Simoncelli, and S.-C. Zhu. On advances in statistical modeling of natural images. *Journal of Mathematical Imaging and Vision*, 18: 17-33, 2003.

- [35] M. Stetter and K. Obermayer. Biology and theory of early vision in mammals. In Szu H., editor, *Brains and Biological Neural Networks*. INNS press, 2000.
- [36] A.L. Yuille, J.M. Coughlan, and D. Kersten. Computational vision: Principles of perceptual inference. Notes for Neural Information Processing Tutorial, Denver, CO., November 1998.
- [37] L. Zhaoping. Optimal sensory encoding. In M.A. Arbib, editor, *The Handbook of Brain Theory and Neural Networks: The Second Edition*, pages 815–819. MIT Press, 2002.
- [38] D. Ziou and S. Tabbone. Edge detection techniques - an overview. *International Journal of Pattern Recognition and Image Analysis*, Vol. 8, No. 4, pp. 537-559, 1998.