

# Natural Image Statistics

# **Computational Imaging and Vision**

---

*Managing Editor*

MAX VIERGEVER

*Utrecht University, The Netherlands*

*Series Editors*

GUNILLA BORGEFORS, *Centre for Image Analysis, SLU, Uppsala, Sweden*

RACHID DERICHE, *INRIA, France*

THOMAS S. HUANG, *University of Illinois, Urbana, U.S.A.*

KATSUSHI IKEUCHI, *Tokyo University, Japan*

TIANZI JIANG, *Institute of Automation, CAS, Beijing*

REINHARD KLETTE, *University of Auckland, New Zealand*

ALES LEONARDIS, *ViCoS, University of Ljubljana, Slovenia*

HEINZ-OTTO PEITGEN, *CeVis, Bremen, Germany*

JOHN K. TSOTSOS, *York University, Canada*

This comprehensive book series embraces state-of-the-art expository works and advanced research monographs on any aspect of this interdisciplinary field.

Topics covered by the series fall in the following four main categories:

- Imaging Systems and Image Processing
- Computer Vision and Image Understanding
- Visualization
- Applications of Imaging Technologies

Only monographs or multi-authored books that have a distinct subject area, that is where each chapter has been invited in order to fulfill this purpose, will be considered for the series.

---

Volume 39

---

For other titles published in this series, go to  
[www.springer.com/series/5754](http://www.springer.com/series/5754)

# Natural Image Statistics

A Probabilistic Approach  
to Early Computational Vision

Written by

**Aapo Hyvärinen**

*University of Helsinki, Finland*

**Jarmo Hurri**

*University of Helsinki, Finland*

and

**Patrik O. Hoyer**

*University of Helsinki, Finland*



Springer

Aapo Hyvärinen  
University of Helsinki  
Dept. Mathematics & Statistics and  
Dept. Computer Science  
Finland

Jarmo Hurri  
University of Helsinki  
Dept. Computer Science  
Finland

Patrik O. Hoyer  
University of Helsinki  
Dept. Computer Science  
Finland

ISSN 1381-6446  
Computational Imaging and Vision  
ISBN 978-1-84882-490-4                    e-ISBN 978-1-84882-491-1  
Springer Dordrecht Heidelberg London New York

British Library Cataloguing in Publication Data  
A catalogue record for this book is available from the British Library

Library of Congress Control Number: 2009923579

AMS Codes: 62H35, 92B20, 91E30, 68U10, 68T05, 94A08, 92C20

©Springer-Verlag London Limited 2009

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act 1988, this publication may only be reproduced, stored or transmitted, in any form or by any means, with the prior permission in writing of the publishers, or in the case of reprographic reproduction in accordance with the terms of licenses issued by the Copyright Licensing Agency. Enquiries concerning reproduction outside those terms should be sent to the publishers.

The use of registered names, trademarks, etc., in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant laws and regulations and therefore free for general use.

The publisher makes no representation, express or implied, with regard to the accuracy of the information contained in this book and cannot accept any legal responsibility or liability for any errors or omissions that may be made.

*Cover design:* WMHDesign GmbH

Printed on acid-free paper

Springer is part of Springer Science+Business Media ([www.springer.com](http://www.springer.com))

# Preface

## Aims and Scope

This book is both an introductory textbook and a research monograph on modeling the statistical structure of natural images. In very simple terms, “natural images” are photographs of the typical environment where we live. In this book, their statistical structure is described using a number of statistical models whose parameters are estimated from image samples.

Our main motivation for exploring natural image statistics is computational modeling of biological visual systems. A theoretical framework which is gaining more and more support considers the properties of the visual system to be reflections of the statistical structure of natural images because of evolutionary adaptation processes. Another motivation for natural image statistics research is in computer science and engineering, where it helps in development of better image processing and computer vision methods.

While research on natural image statistics has been growing rapidly since the mid-1990s, no attempt has been made to cover the field in a single book, providing a unified view of the different models and approaches. This book attempts to do just that. Furthermore, our aim is to provide an accessible introduction to the field for students in related disciplines.

However, not all aspects of such a large field of study can be completely covered in a single book, so we have had to make some choices. Basically, we concentrate on the neural modeling approaches at the expense of engineering applications. Furthermore, those topics on which the authors themselves have been doing research are inevitably given more emphasis.

## Targeted Audience and Prerequisites

The book is targeted for advanced undergraduate students, graduate students and researchers in vision science, computational neuroscience, computer vision, and image processing. It can also be read as an introduction to the area by people with a background in mathematical disciplines (mathematics, statistics, theoretical physics).

Due to the multi-disciplinary nature of the subject, the book has been written so as to be accessible to an audience coming from very different backgrounds such as psychology, computer science, electrical engineering, neurobiology, mathematics, statistics, and physics. Therefore, we have attempted to reduce the prerequisites to a minimum. The main thing needed are basic mathematical skills as taught in introductory university-level mathematics courses. In particular, the reader is assumed to

know the basics of

- univariate calculus (e.g. one-dimensional derivatives and integrals)
- linear algebra (e.g. inverse matrix, orthogonality)
- probability and statistics (e.g. expectation, probability density function, variance, covariance)

To help readers with a modest mathematical background, a crash course on linear algebra is offered at Chap. 19, and Chap. 4 reviews probability theory and statistics on a rather elementary level.

No previous knowledge of neuroscience or vision science is necessary for reading this book. All the necessary background on the visual system is given in Chap. 3, and an introduction to some basic image processing methods is given in Chap. 2.

## Structure of the Book and Its Use as a Textbook

This book is a hybrid of a monograph and an advanced graduate textbook. It starts with background material which is rather classic, whereas the latter parts of the book consider very recent work with many open problems. The material in the middle is quite recent but relatively established.

The book is divided into the following parts:

**Introduction** which explains the basic setting and motivation.

**Part I** which consists of background chapters. This is mainly classic material found in many textbooks in statistics, neuroscience, and signal processing. However, here it has been carefully selected to ensure that the reader has the right background for the main part of the book.

**Part II** starts the main topic, considering the most basic models for natural image statistics. These models are based on the statistics of linear features, i.e. linear combinations of image pixel values.

**Part III** considers more sophisticated models of natural image statistics, in which dependencies (interactions) of linear features are considered, which is related to computing non-linear features.

**Part IV** applies the models already introduced to new kinds of data: color images, stereo images, and image sequences (video). Some new models on the temporal structure of sequences are also introduced.

**Part V** consists of a concluding chapter. It provides a short overview of the book and discusses open questions as well as alternative approaches to image modeling.

**Part VI** consists of mathematical chapters which are provided as a kind of an appendix. Chapter 18 is a rather independent chapter on optimization theory. Chapter 19 is background material which the reader is actually supposed to know; it is provided here as a reminder. Chapters 20 and 21 provide sophisticated supplementary mathematical material for readers with such interests.

Dependencies of the parts are rather simple. When the book is used as a textbook, **all readers should start by reading the first seven chapters** in the order they are

given (i.e. Introduction, Part I, and Part II except for the last chapter), unless the reader is already familiar with some of the material. After that, it is possible to jump to later chapters in almost any order, except for the following:

- Chapter 10 requires Chap. 9, and Chap. 11 requires Chaps. 9 and 10.
- Chapter 14 requires Sect. 13.1.

Some of the sections are marked with an asterisk \*, which means that they are more sophisticated material which can be skipped without interrupting the flow of ideas.

An introductory course on natural image statistics can be simply constructed by going through the first  $n$  chapters of the book, where  $n$  would typically be between 7 and 17, depending on the amount of time available.

## Referencing and Exercises

To keep the text readable and suitable for a textbook, the first 11 chapters do not include references in the main text. References are given in a separate section at the end of the chapter. In the latter chapters, the nature of the material requires that references are given in the text, so the style changes to a more scholarly one. Likewise, mathematical exercises and computer assignments are given for the first 10 chapters.

## Code for Reproducing Experiments

For pedagogical purposes as well as to ensure the reproducibility of the experiments, the Matlab™ code for producing most of the experiments in the first 11 chapters, and some in Chap. 13, is distributed on the Internet at

[www.naturalimagestatistics.net](http://www.naturalimagestatistics.net)

This web site will also include other related material.

## Acknowledgements

We would like to thank Michael Gutmann, Asun Vicente, and Jussi Lindgren for detailed comments on the manuscript. We have also greatly benefited from discussions with Bruno Olshausen, Eero Simoncelli, Geoffrey Hinton, David Field, Peter Dayan, David Donoho, Pentti Laurinen, Jussi Saarinen, Simo Vanni, and many others. We are also very grateful to Dario Ringach for providing the reverse correlation results in Fig. 3.7. During the writing process, the authors were funded by the University of Helsinki (Department of Computer Science and Department of Mathematics and Statistics), the Helsinki Institute for Information Technology, and the Academy of Finland.

Helsinki

Aapo Hyvärinen  
Jarmo Hurri  
Patrik Hoyer

# Contents

<b>1</b>	<b>Introduction</b>	1
1.1	What this Book Is All About	1
1.2	What Is Vision?	2
1.3	The Magic of Your Visual System	3
1.4	Importance of Prior Information	7
1.4.1	Ecological Adaptation Provides Prior Information	7
1.4.2	Generative Models and Latent Quantities	8
1.4.3	Projection onto the Retina Loses Information	9
1.4.4	Bayesian Inference and Priors	9
1.5	Natural Images	10
1.5.1	The Image Space	10
1.5.2	Definition of Natural Images	11
1.6	Redundancy and Information	13
1.6.1	Information Theory and Image Coding	13
1.6.2	Redundancy Reduction and Neural Coding	14
1.7	Statistical Modeling of the Visual System	15
1.7.1	Connecting Information Theory and Bayesian Inference	15
1.7.2	Normative vs. Descriptive Modeling of Visual System	15
1.7.3	Toward Predictive Theoretical Neuroscience	16
1.8	Features and Statistical Models of Natural Images	17
1.8.1	Image Representations and Features	17
1.8.2	Statistics of Features	18
1.8.3	From Features to Statistical Models	19
1.9	The Statistical–Ecological Approach Recapitulated	20
1.10	References	21

## Part I Background

<b>2</b>	<b>Linear Filters and Frequency Analysis</b>	25
2.1	Linear Filtering	25
2.1.1	Definition	25
2.1.2	Impulse Response and Convolution	28
2.2	Frequency-Based Representation	29
2.2.1	Motivation	29
2.2.2	Representation in One and Two Dimensions	29
2.2.3	Frequency-Based Representation and Linear Filtering	34
2.2.4	Computation and Mathematical Details	37
2.3	Representation Using Linear Basis	38
2.3.1	Basic Idea	38
2.3.2	Frequency-Based Representation as a Basis	40

2.4	Space-Frequency Analysis . . . . .	41
2.4.1	Introduction . . . . .	41
2.4.2	Space-Frequency Analysis and Gabor Filters . . . . .	43
2.4.3	Spatial Localization vs. Spectral Accuracy . . . . .	46
2.5	References . . . . .	48
2.6	Exercises . . . . .	48
<b>3</b>	<b>Outline of the Visual System . . . . .</b>	<b>51</b>
3.1	Neurons and Firing Rates . . . . .	51
3.2	From the Eye to the Cortex . . . . .	53
3.3	Linear Models of Visual Neurons . . . . .	54
3.3.1	Responses to Visual Stimulation . . . . .	54
3.3.2	Simple Cells and Linear Models . . . . .	56
3.3.3	Gabor Models and Selectivities of Simple Cells . . . . .	57
3.3.4	Frequency Channels . . . . .	58
3.4	Non-linear Models of Visual Neurons . . . . .	59
3.4.1	Non-linearities in Simple-Cell Responses . . . . .	59
3.4.2	Complex Cells and Energy Models . . . . .	61
3.5	Interactions between Visual Neurons . . . . .	62
3.6	Topographic Organization . . . . .	64
3.7	Processing after the Primary Visual Cortex . . . . .	64
3.8	References . . . . .	65
3.9	Exercises . . . . .	65
<b>4</b>	<b>Multivariate Probability and Statistics . . . . .</b>	<b>67</b>
4.1	Natural Images Patches as Random Vectors . . . . .	67
4.2	Multivariate Probability Distributions . . . . .	68
4.2.1	Notation and Motivation . . . . .	68
4.2.2	Probability Density Function . . . . .	69
4.3	Marginal and Joint Probabilities . . . . .	70
4.4	Conditional Probabilities . . . . .	73
4.5	Independence . . . . .	75
4.6	Expectation and Covariance . . . . .	77
4.6.1	Expectation . . . . .	77
4.6.2	Variance and Covariance in One Dimension . . . . .	78
4.6.3	Covariance Matrix . . . . .	78
4.6.4	Independence and Covariances . . . . .	79
4.7	Bayesian Inference . . . . .	81
4.7.1	Motivating Example . . . . .	81
4.7.2	Bayes' Rule . . . . .	83
4.7.3	Non-informative Priors . . . . .	83
4.7.4	Bayesian Inference as an Incremental Learning Process . . . . .	84
4.8	Parameter Estimation and Likelihood . . . . .	86
4.8.1	Models, Estimation, and Samples . . . . .	86
4.8.2	Maximum Likelihood and Maximum a Posteriori . . . . .	87
4.8.3	Prior and Large Samples . . . . .	89

4.9 References . . . . .	89
4.10 Exercises . . . . .	89

## Part II Statistics of Linear Features

<b>5 Principal Components and Whitening . . . . .</b>	<b>93</b>
5.1 DC Component or Mean Grey-Scale Value . . . . .	93
5.2 Principal Component Analysis . . . . .	94
5.2.1 A Basic Dependency of Pixels in Natural Images . . . . .	94
5.2.2 Learning One Feature by Maximization of Variance . . . . .	96
5.2.3 Learning Many Features by PCA . . . . .	98
5.2.4 Computational Implementation of PCA . . . . .	101
5.2.5 The Implications of Translation-Invariance . . . . .	102
5.3 PCA as a Preprocessing Tool . . . . .	103
5.3.1 Dimension Reduction by PCA . . . . .	103
5.3.2 Whitening by PCA . . . . .	104
5.3.3 Anti-aliasing by PCA . . . . .	106
5.4 Canonical Preprocessing Used in This Book . . . . .	109
5.5 Gaussianity as the Basis for PCA . . . . .	109
5.5.1 The Probability Model Related to PCA . . . . .	109
5.5.2 PCA as a Generative Model . . . . .	110
5.5.3 Image Synthesis Results . . . . .	111
5.6 Power Spectrum of Natural Images . . . . .	111
5.6.1 The $1/f$ Fourier Amplitude or $1/f^2$ Power Spectrum . . . . .	111
5.6.2 Connection between Power Spectrum and Covariances . . . . .	113
5.6.3 Relative Importance of Amplitude and Phase . . . . .	114
5.7 Anisotropy in Natural Images . . . . .	115
5.8 Mathematics of Principal Component Analysis* . . . . .	116
5.8.1 Eigenvalue Decomposition of the Covariance Matrix . . . . .	117
5.8.2 Eigenvectors and Translation-Invariance . . . . .	119
5.9 Decorrelation Models of Retina and LGN * . . . . .	120
5.9.1 Whitening and Redundancy Reduction . . . . .	120
5.9.2 Patch-Based Decorrelation . . . . .	121
5.9.3 Filter-Based Decorrelation . . . . .	124
5.10 Concluding Remarks and References . . . . .	128
5.11 Exercises . . . . .	129
<b>6 Sparse Coding and Simple Cells . . . . .</b>	<b>131</b>
6.1 Definition of Sparseness . . . . .	131
6.2 Learning One Feature by Maximization of Sparseness . . . . .	132
6.2.1 Measuring Sparseness: General Framework . . . . .	133
6.2.2 Measuring Sparseness Using Kurtosis . . . . .	133
6.2.3 Measuring Sparseness Using Convex Functions of Square .	134
6.2.4 The Case of Canonically Preprocessed Data . . . . .	138
6.2.5 One Feature Learned from Natural Images . . . . .	138

6.3	Learning Many Features by Maximization of Sparseness . . . . .	139
6.3.1	Deflationary Decorrelation . . . . .	140
6.3.2	Symmetric Decorrelation . . . . .	141
6.3.3	Sparseness of Feature vs. Sparseness of Representation . .	141
6.4	Sparse Coding Features for Natural Images . . . . .	143
6.4.1	Full Set of Features . . . . .	143
6.4.2	Analysis of Tuning Properties . . . . .	144
6.5	How Is Sparseness Useful? . . . . .	147
6.5.1	Bayesian Modeling . . . . .	147
6.5.2	Neural Modeling . . . . .	148
6.5.3	Metabolic Economy . . . . .	148
6.6	Concluding Remarks and References . . . . .	148
6.7	Exercises . . . . .	149
<b>7</b>	<b>Independent Component Analysis . . . . .</b>	<b>151</b>
7.1	Limitations of the Sparse Coding Approach . . . . .	151
7.2	Definition of ICA . . . . .	152
7.2.1	Independence . . . . .	152
7.2.2	Generative Model . . . . .	152
7.2.3	Model for Preprocessed Data . . . . .	154
7.3	Insufficiency of Second-Order Information . . . . .	154
7.3.1	Why Whitening Does Not Find Independent Components .	154
7.3.2	Why Components Have to Be Non-Gaussian . . . . .	156
7.4	The Probability Density Defined by ICA . . . . .	158
7.5	Maximum Likelihood Estimation in ICA . . . . .	159
7.6	Results on Natural Images . . . . .	160
7.6.1	Estimation of Features . . . . .	160
7.6.2	Image Synthesis Using ICA . . . . .	160
7.7	Connection to Maximization of Sparseness . . . . .	161
7.7.1	Likelihood as a Measure of Sparseness . . . . .	161
7.7.2	Optimal Sparseness Measures . . . . .	163
7.8	Why Are Independent Components Sparse? . . . . .	166
7.8.1	Different Forms of Non-Gaussianity . . . . .	167
7.8.2	Non-Gaussianity in Natural Images . . . . .	167
7.8.3	Why Is Sparseness Dominant? . . . . .	168
7.9	General ICA as Maximization of Non-Gaussianity . . . . .	168
7.9.1	Central Limit Theorem . . . . .	169
7.9.2	“Non-Gaussian Is Independent” . . . . .	169
7.9.3	Sparse Coding as a Special Case of ICA . . . . .	170
7.10	Receptive Fields vs. Feature Vectors . . . . .	171
7.11	Problem of Inversion of Preprocessing . . . . .	172
7.12	Frequency Channels and ICA . . . . .	173
7.13	Concluding Remarks and References . . . . .	173
7.14	Exercises . . . . .	174

<b>8 Information-Theoretic Interpretations . . . . .</b>	177
8.1 Basic Motivation for Information Theory . . . . .	177
8.1.1 Compression . . . . .	177
8.1.2 Transmission . . . . .	178
8.2 Entropy as a Measure of Uncertainty . . . . .	179
8.2.1 Definition of Entropy . . . . .	179
8.2.2 Entropy as Minimum Coding Length . . . . .	180
8.2.3 Redundancy . . . . .	181
8.2.4 Differential Entropy . . . . .	182
8.2.5 Maximum Entropy . . . . .	183
8.3 Mutual Information . . . . .	184
8.4 Minimum Entropy Coding of Natural Images . . . . .	185
8.4.1 Image Compression and Sparse Coding . . . . .	185
8.4.2 Mutual Information and Sparse Coding . . . . .	187
8.4.3 Minimum Entropy Coding in the Cortex . . . . .	187
8.5 Information Transmission in the Nervous System . . . . .	188
8.5.1 Definition of Information Flow and Infomax . . . . .	188
8.5.2 Basic Infomax with Linear Neurons . . . . .	188
8.5.3 Infomax with Non-linear Neurons . . . . .	189
8.5.4 Infomax with Non-constant Noise Variance . . . . .	190
8.6 Caveats in Application of Information Theory . . . . .	193
8.7 Concluding Remarks and References . . . . .	195
8.8 Exercises . . . . .	195

### Part III Nonlinear Features and Dependency of Linear Features

<b>9 Energy Correlation of Linear Features and Normalization . . . . .</b>	199
9.1 Why Estimated Independent Components Are Not Independent . . . . .	199
9.1.1 Estimates vs. Theoretical Components . . . . .	199
9.1.2 Counting the Number of Free Parameters . . . . .	200
9.2 Correlations of Squares of Components in Natural Images . . . . .	201
9.3 Modeling Using a Variance Variable . . . . .	201
9.4 Normalization of Variance and Contrast Gain Control . . . . .	203
9.5 Physical and Neurophysiological Interpretations . . . . .	205
9.5.1 Canceling the Effect of Changing Lighting Conditions . . . . .	205
9.5.2 Uniform Surfaces . . . . .	206
9.5.3 Saturation of Cell Responses . . . . .	206
9.6 Effect of Normalization on ICA . . . . .	207
9.7 Concluding Remarks and References . . . . .	210
9.8 Exercises . . . . .	211
<b>10 Energy Detectors and Complex Cells . . . . .</b>	213
10.1 Subspace Model of Invariant Features . . . . .	213
10.1.1 Why Linear Features Are Insufficient . . . . .	213
10.1.2 Subspaces or Groups of Linear Features . . . . .	213
10.1.3 Energy Model of Feature Detection . . . . .	214

10.2	Maximizing Sparseness in the Energy Model . . . . .	216
10.2.1	Definition of Sparseness of Output . . . . .	216
10.2.2	One Feature Learned from Natural Images . . . . .	217
10.3	Model of Independent Subspace Analysis . . . . .	219
10.4	Dependency as Energy Correlation . . . . .	220
10.4.1	Why Energy Correlations Are Related to Sparseness . . . . .	220
10.4.2	Spherical Symmetry and Changing Variance . . . . .	221
10.4.3	Correlation of Squares and Convexity of Non-linearity . . . . .	222
10.5	Connection to Contrast Gain Control . . . . .	223
10.6	ISA as a Non-linear Version of ICA . . . . .	224
10.7	Results on Natural Images . . . . .	225
10.7.1	Emergence of Invariance to Phase . . . . .	225
10.7.2	The Importance of Being Invariant . . . . .	230
10.7.3	Grouping of Dependencies . . . . .	232
10.7.4	Superiority of the Model over ICA . . . . .	232
10.8	Analysis of Convexity and Energy Correlations* . . . . .	234
10.8.1	Variance Variable Model Gives Convex $h$ . . . . .	234
10.8.2	Convex $h$ Typically Implies Positive Energy Correlations . . . . .	235
10.9	Concluding Remarks and References . . . . .	236
10.10	Exercises . . . . .	236
<b>11</b>	<b>Energy Correlations and Topographic Organization</b> . . . . .	239
11.1	Topography in the Cortex . . . . .	239
11.2	Modeling Topography by Statistical Dependence . . . . .	240
11.2.1	Topographic Grid . . . . .	240
11.2.2	Defining Topography by Statistical Dependencies . . . . .	240
11.3	Definition of Topographic ICA . . . . .	242
11.4	Connection to Independent Subspaces and Invariant Features . . . . .	243
11.5	Utility of Topography . . . . .	244
11.6	Estimation of Topographic ICA . . . . .	245
11.7	Topographic ICA of Natural Images . . . . .	246
11.7.1	Emergence of V1-like Topography . . . . .	246
11.7.2	Comparison with Other Models . . . . .	253
11.8	Learning Both Layers in a Two-Layer Model * . . . . .	253
11.8.1	Generative vs. Energy-Based Approach . . . . .	253
11.8.2	Definition of the Generative Model . . . . .	254
11.8.3	Basic Properties of the Generative Model . . . . .	255
11.8.4	Estimation of the Generative Model . . . . .	256
11.8.5	Energy-Based Two-Layer Models . . . . .	259
11.9	Concluding Remarks and References . . . . .	260
<b>12</b>	<b>Dependencies of Energy Detectors: Beyond V1</b> . . . . .	263
12.1	Predictive Modeling of Extrastriate Cortex . . . . .	263
12.2	Simulation of V1 by a Fixed Two-Layer Model . . . . .	263
12.3	Learning the Third Layer by Another ICA Model . . . . .	265

12.4	Methods for Analyzing Higher-Order Components . . . . .	266
12.5	Results on Natural Images . . . . .	268
12.5.1	Emergence of Collinear Contour Units . . . . .	268
12.5.2	Emergence of Pooling over Frequencies . . . . .	269
12.6	Discussion of Results . . . . .	273
12.6.1	Why Coding of Contours? . . . . .	273
12.6.2	Frequency Channels and Edges . . . . .	274
12.6.3	Toward Predictive Modeling . . . . .	274
12.6.4	References and Related Work . . . . .	275
12.7	Conclusion . . . . .	276
<b>13</b>	<b>Overcomplete and Non-negative Models</b> . . . . .	277
13.1	Overcomplete Bases . . . . .	277
13.1.1	Motivation . . . . .	277
13.1.2	Definition of Generative Model . . . . .	278
13.1.3	Nonlinear Computation of the Basis Coefficients . . . . .	279
13.1.4	Estimation of the Basis . . . . .	281
13.1.5	Approach Using Energy-Based Models . . . . .	282
13.1.6	Results on Natural Images . . . . .	285
13.1.7	Markov Random Field Models *	285
13.2	Non-negative Models . . . . .	288
13.2.1	Motivation . . . . .	288
13.2.2	Definition . . . . .	288
13.2.3	Adding Sparseness Constraints . . . . .	290
13.3	Conclusion . . . . .	293
<b>14</b>	<b>Lateral Interactions and Feedback</b> . . . . .	295
14.1	Feedback as Bayesian Inference . . . . .	295
14.1.1	Example: Contour Integrator Units . . . . .	296
14.1.2	Thresholding (Shrinkage) of a Sparse Code . . . . .	298
14.1.3	Categorization and Top-Down Feedback . . . . .	302
14.2	Overcomplete Basis and End-stopping . . . . .	302
14.3	Predictive Coding . . . . .	304
14.4	Conclusion . . . . .	305
<b>Part IV Time, Color, and Stereo</b>		
<b>15</b>	<b>Color and Stereo Images</b> . . . . .	309
15.1	Color Image Experiments . . . . .	309
15.1.1	Choice of Data . . . . .	309
15.1.2	Preprocessing and PCA . . . . .	310
15.1.3	ICA Results and Discussion . . . . .	313
15.2	Stereo Image Experiments . . . . .	315
15.2.1	Choice of Data . . . . .	315
15.2.2	Preprocessing and PCA . . . . .	316
15.2.3	ICA Results and Discussion . . . . .	317

15.3	Further References . . . . .	322
15.3.1	Color and Stereo Images . . . . .	322
15.3.2	Other Modalities, Including Audition . . . . .	323
15.4	Conclusion . . . . .	323
<b>16</b>	<b>Temporal Sequences of Natural Images . . . . .</b>	<b>325</b>
16.1	Natural Image Sequences and Spatiotemporal Filtering . . . . .	325
16.2	Temporal and Spatiotemporal Receptive Fields . . . . .	326
16.3	Second-Order Statistics . . . . .	328
16.3.1	Average Spatiotemporal Power Spectrum . . . . .	328
16.3.2	The Temporally Decorrelating Filter . . . . .	332
16.4	Sparse Coding and ICA of Natural Image Sequences . . . . .	333
16.5	Temporal Coherence in Spatial Features . . . . .	336
16.5.1	Temporal Coherence and Invariant Representation . . . . .	336
16.5.2	Quantifying Temporal Coherence . . . . .	337
16.5.3	Interpretation as Generative Model *	338
16.5.4	Experiments on Natural Image Sequences . . . . .	339
16.5.5	Why Gabor-Like Features Maximize Temporal Coherence .	341
16.5.6	Control Experiments . . . . .	344
16.6	Spatiotemporal Energy Correlations in Linear Features . . . . .	345
16.6.1	Definition of the Model . . . . .	345
16.6.2	Estimation of the Model . . . . .	347
16.6.3	Experiments on Natural Images . . . . .	348
16.6.4	Intuitive Explanation of Results . . . . .	350
16.7	Unifying Model of Spatiotemporal Dependencies . . . . .	352
16.8	Features with Minimal Average Temporal Change . . . . .	354
16.8.1	Slow Feature Analysis . . . . .	354
16.8.2	Quadratic Slow Feature Analysis . . . . .	357
16.8.3	Sparse Slow Feature Analysis . . . . .	359
16.9	Conclusion . . . . .	361

## Part V Conclusion

<b>17</b>	<b>Conclusion and Future Prospects . . . . .</b>	<b>365</b>
17.1	Short Overview . . . . .	365
17.2	Open, or Frequently Asked, Questions . . . . .	367
17.2.1	What Is the Real Learning Principle in the Brain? . . . . .	367
17.2.2	Nature vs. Nurture . . . . .	368
17.2.3	How to Model Whole Images . . . . .	369
17.2.4	Are There Clear-Cut Cell Types? . . . . .	369
17.2.5	How Far Can We Go? . . . . .	371
17.3	Other Mathematical Models of Images . . . . .	371
17.3.1	Scaling Laws . . . . .	372
17.3.2	Wavelet Theory . . . . .	372
17.3.3	Physically Inspired Models . . . . .	373
17.4	Future Work . . . . .	374

**Part VI Appendix: Supplementary Mathematical Tools**

<b>18 Optimization Theory and Algorithms . . . . .</b>	377
18.1 Levels of Modeling . . . . .	377
18.2 Gradient Method . . . . .	378
18.2.1 Definition and Meaning of Gradient . . . . .	378
18.2.2 Gradient and Optimization . . . . .	380
18.2.3 Optimization of Function of Matrix . . . . .	381
18.2.4 Constrained Optimization . . . . .	381
18.3 Global and Local Maxima . . . . .	383
18.4 Hebb's Rule and Gradient Methods . . . . .	384
18.4.1 Hebb's Rule . . . . .	384
18.4.2 Hebb's Rule and Optimization . . . . .	385
18.4.3 Stochastic Gradient Methods . . . . .	386
18.4.4 Role of the Hebbian Non-linearity . . . . .	387
18.4.5 Receptive Fields vs. Synaptic Strengths . . . . .	388
18.4.6 The Problem of Feedback . . . . .	388
18.5 Optimization in Topographic ICA *	389
18.6 Beyond Basic Gradient Methods *	390
18.6.1 Newton's Method . . . . .	391
18.6.2 Conjugate Gradient Methods . . . . .	393
18.7 FastICA, a Fixed-Point Algorithm for ICA . . . . .	394
18.7.1 The FastICA Algorithm . . . . .	394
18.7.2 Choice of the FastICA Non-linearity . . . . .	395
18.7.3 Mathematics of FastICA *	395
<b>19 Crash Course on Linear Algebra . . . . .</b>	399
19.1 Vectors . . . . .	399
19.2 Linear Transformations . . . . .	400
19.3 Matrices . . . . .	401
19.4 Determinant . . . . .	402
19.5 Inverse . . . . .	402
19.6 Basis Representations . . . . .	403
19.7 Orthogonality . . . . .	404
19.8 Pseudo-Inverse *	405
<b>20 The Discrete Fourier Transform . . . . .</b>	407
20.1 Linear Shift-Invariant Systems . . . . .	407
20.2 One-Dimensional Discrete Fourier Transform . . . . .	408
20.2.1 Euler's Formula . . . . .	408
20.2.2 Representation in Complex Exponentials . . . . .	408
20.2.3 The Discrete Fourier Transform and Its Inverse . . . . .	411
20.3 Two- and Three-Dimensional Discrete Fourier Transforms . . . . .	417
<b>21 Estimation of Non-normalized Statistical Models . . . . .</b>	419
21.1 Non-normalized Statistical Models . . . . .	419

21.2 Estimation by Score Matching . . . . .	420
21.3 Example 1: Multivariate Gaussian Density . . . . .	422
21.4 Example 2: Estimation of Basic ICA Model . . . . .	424
21.5 Example 3: Estimation of an Overcomplete ICA Model . . . . .	425
21.6 Conclusion . . . . .	425
<b>References</b> . . . . .	427
<b>Index</b> . . . . .	441

# Abbreviations

DFT	discrete Fourier transform
FFT	fast Fourier transform
ICA	independent component analysis
ISA	independent subspace analysis
LGN	lateral geniculate nucleus
MAP	maximum a posteriori
MRF	Markov random field
NMF	non-negative matrix factorization
PCA	principal component analysis
RF	receptive field
RGB	red–green–blue
V1	primary visual cortex
V2, V3, ...	other visual cortical areas