Contents

1 Introduction **1**
- 1.1 Goal of Vision: Understanding Scene and Events . . . 1
- 1.2 Knowledge Representation: Sets, Concepts and Models . . . 3
- 1.3 Unification: Regimes of Representations . . . 5
- 1.4 Conceptualization: Sets and Statistical Models . . . 6

2 Statistics of Natural Images **9**
- 2.1 Image Space and Distribution . . . 9
- 2.2 Information and Encoding . . . 11
- 2.3 Image Statistics and Power Law . . . 13
- 2.4 Kurtosis and Sparsity . . . 15
- 2.5 Scale Invariance . . . 18

3 Textures **21**
- 3.1 The Julesz Quest . . . 21
- 3.2 Clique-Based Gibbs Models & MRF . . . 23
  - 3.2.1 Markov Random Fields (MRF) . . . 23
  - 3.2.2 Ising & Potts Model . . . 24
  - 3.2.3 Gaussian Markov Random Fields (GMRF) . . . 26
- 3.3 Filters for Early Vision . . . 28
  - 3.3.1 Correlation & Convolution . . . 29
  - 3.3.2 Edge Detection Filters . . . 30
  - 3.3.3 Gaussian Filter . . . 31
  - 3.3.4 Derivative of Gaussian & Laplacian of Gaussian Filter . . . 32
  - 3.3.5 Gabor Filter . . . 34
- 3.4 The FRAME Model . . . 35
  - 3.4.1 Intuition and the Big Picture . . . 35
  - 3.4.2 Deriving the FRAME model . . . 37
- 3.5 The Texture Ensemble . . . 45
  - 3.5.1 Ensembles in Statistical Physics . . . 46
  - 3.5.2 Texture Ensemble . . . 47
  - 3.5.3 Type Theory and Entropy Rate Functions . . . 49
  - 3.5.4 Equivalence of FRAME and Julesz Ensemble . . . 52
- 3.6 Deriving Partial Differential Equations from the MRF and FRAME models . . . 55
  - 3.6.1 Turing Diffusion-Reaction . . . 55
  - 3.6.2 Heat Diffusion . . . 56

3.6.3 Anisotropic Diffusion . . . 56
3.6.4 GRADE: Gibbs Reaction And Diffusion Equations . . . 57
3.6.5 Properties of GRADE . . . 59
3.7 Conclusion . . . 60

4 Textons **63**
4.1 Julesz's Confusion between Textures and Textons . . . 63
4.2 Sparse Coding . . . 64
4.2.1 Image Representation . . . 64
4.2.2 Olshausen-Field model . . . 65
4.2.3 A three-level generative model . . . 67
4.3 Active basis model . . . 72
4.3.1 Olshausen-Field model for sparse coding . . . 72
4.3.2 Active basis model for shared sparse coding of aligned image patches . . . 73
4.3.3 Prototype algorithm . . . 75
4.3.4 Statistical modeling . . . 76
4.3.5 Shared matching pursuit . . . 77
4.4 Compositional Sparse Coding . . . 79
4.4.1 Sparsity and Composition . . . 79
4.4.2 Compositional sparse coding model . . . 81
4.5 The Sparse FRAME Model . . . 84
4.5.1 Dense FRAME . . . 85
4.5.2 Sparse representation . . . 85
4.5.3 Maximum likelihood learning . . . 85
4.5.4 Generative boosting . . . 86
4.5.5 Sparse model . . . 87

5 Primal Sketch: Integrating Textures and Textons **91**
5.1 Marr's Conjecture on Primal Sketch . . . 91
5.2 The Two Layer Model . . . 92
5.2.1 Structure domain . . . 93
5.2.2 Texture domain . . . 95
5.2.3 Integrated model . . . 95
5.3 Hybrid Image Templates . . . 96
5.3.1 Representation . . . 97
5.3.2 Prototypes, ϵ - balls, and saturation function . . . 98
5.3.3 Projecting image patches to 1D responses . . . 99
5.3.4 Template pursuit by information projection . . . 101
5.4 Relations between Primal Sketch and the HoG and SIFT Representations . . . 103

6 2.1D Sketch and Layered Representation **105**
6.1 Problem Formulation . . . 106
6.2 The Variational Formulation by Nitzberg and Mumford . . . 107
6.2.1 The Energy Functional . . . 107
6.2.2 The Euler Elastica for Completing Occluded Curves . . . 108
6.3 The Mixed Markov Random Field Formulation . . . 108
6.3.1 Definition of W_{2D} and $W_{2.1D}$ . . . 109

6.3.2 The mixed MRF and Its Graphical Representation . . . 111
6.3.3 Bayesian formulation . . . 112
6.3.4 Experiments . . . 113
6.4 The 2.1D Sketch with Layered Regions and Curves . . . 114
6.4.1 Generative models and Bayesian formulation . . . 114
6.4.2 Bayesian formulation for probabilistic inference . . . 118
6.4.3 Experiments . . . 119

7 2.5D Sketch and Depth Maps 123
7.1 Marr's Definition . . . 123
7.2 2.5D Sketch from Primal Sketch — Shape from Stereo . . . 124
7.2.1 Two Layer Representation . . . 126
7.2.2 The inference algorithm . . . 133
7.2.3 Example results . . . 136
7.3 2.5D Sketch from Primal Sketch — Shape from Shading . . . 136
7.3.1 Overview of the Two-Layer Generation Model . . . 137
7.3.2 Results . . . 138
7.4 2.5D Sketch from Direct Estimation . . . 139
7.4.1 Dataset . . . 139
7.4.2 Model . . . 141
7.4.3 Results . . . 143

8 Learning by Information Projection 145
8.1 Information projection . . . 145
8.1.1 Orthogonality and duality . . . 146
8.1.2 Maximum Likelihood Implementation . . . 147
8.1.3 The Minimax Learning Framework . . . 148
8.1.4 Model Pursuit Strategies . . . 149
8.2 A Unifying View . . . 150
8.2.1 Relation to Discriminative Learning . . . 150
8.2.2 Learning FRAME . . . 150
8.2.3 Learning Shape Patterns . . . 151

9 Information Scaling and Regimes of Models 155
9.1 Image Scaling . . . 155
9.2 Perceptual Entropy . . . 163
9.3 A Continuous Spectrum . . . 166
9.4 Two Coding Schemes . . . 166
9.5 Perceptual Scale Space . . . 168
9.6 Perceptibility, Metastability, and the Energy Landscape . . . 169

10 Image Models with Multilayer Neural Networks 175
10.1 Deep FRAME . . . 175
10.1.1 From FRAME to deep FRAME . . . 177
10.1.2 ConvNet filters . . . 182
10.1.3 FRAME with ConvNet filters . . . 183
10.1.4 Learning and sampling . . . 183

10.1.5 Learning a new layer of filters . . . 185
10.1.6 Deep convolutional energy-based model . . . 187
10.1.7 Hopfield auto-encoder . . . 189
10.1.8 Contrastive Divergence and Hopfield Memory . . . 190
10.1.9 Multigrid sampling and modeling . . . 191
10.1.10 Adversarial interpretation . . . 194
10.1.11 Short-run MCMC . . . 194
10.2 Generator Network . . . 200
10.2.1 Factor analysis . . . 200
10.2.2 Non-linear factor analysis . . . 201
10.2.3 Learning by alternating back-propagation . . . 202
10.2.4 EM, density mapping, and density shifting . . . 205
10.2.5 Extracting appearance and geometry, nonlinear generalization of the AAM model . . . 207
10.2.6 Dynamic generator model . . . 208
10.2.7 Unsupervised clustering and semi-supervised classification . . . 213
10.2.8 Short-run inference dynamics . . . 214

11 A Tale of Three Families: Discriminative, Generative and Descriptive Models 219
11.1 Introduction . . . 219
11.2 Non-hierarchical linear forms of the three families . . . 220
11.2.1 Discriminative models . . . 220
11.2.2 Descriptive model . . . 221
11.2.3 Generative models . . . 224
11.3 Interactions between different families . . . 228
11.3.1 Discriminative learning of descriptive model . . . 228
11.3.2 DDMCMC: integration of discriminative and generative models . . . 228
11.4 Hierarchical forms of the three families . . . 229
11.4.1 Recent developments . . . 229
11.4.2 Discriminative models by convolutional neural networks . . . 230
11.4.3 Descriptive models . . . 232
11.4.4 Introspective learning . . . 233
11.4.5 Generative models . . . 234
11.5 Variational, adversarial and cooperative learning . . . 237
11.5.1 Variational auto-encoder . . . 237
11.5.2 Adversarial contrastive divergence . . . 239
11.5.3 Integrating variational and adversarial learning by divergence triangle . . . 240
11.5.4 Cooperative learning . . . 243
11.5.5 Conditional learning via fast thinking initializer and slow thinking solver . . . 245
11.6 Discussion . . . 246

Index 247

Bibliography 251