

Contents

	About the Author	<i>ix</i>
	Acknowledgments	<i>xi</i>
	Preface	<i>xiii</i>
	Acronyms	<i>xv</i>
1	Introduction	<i>1</i>
1.1	What Is Cognitive Vision	<i>2</i>
1.2	Computational Approaches for Cognitive Vision	<i>3</i>
1.3	A Brief Review of Human Vision System	<i>4</i>
1.4	Perception and Cognition	<i>6</i>
1.5	Organization of the Book	<i>7</i>
2	Early Vision	<i>9</i>
2.1	Feature Integration Theory	<i>9</i>
2.2	Structure of Human Eye	<i>10</i>
2.3	Lateral Inhibition	<i>13</i>
2.4	Convolution: Detection of Edges and Orientations	<i>14</i>
2.5	Color and Texture Perception	<i>17</i>
2.6	Motion Perception	<i>19</i>
2.6.1	Intensity-Based Approach	<i>19</i>
2.6.2	Token-Based Approach	<i>20</i>
2.7	Peripheral Vision	<i>21</i>
2.8	Conclusion	<i>24</i>
3	Bayesian Reasoning for Perception and Cognition	<i>25</i>
3.1	Reasoning Paradigms	<i>26</i>
3.2	Natural Scene Statistics	<i>27</i>
3.3	Bayesian Framework of Reasoning	<i>28</i>
3.4	Bayesian Networks	<i>32</i>

3.5	Dynamic Bayesian Networks	34
3.6	Parameter Estimation	36
3.7	On Complexity of Models and Bayesian Inference	38
3.8	Hierarchical Bayesian Models	39
3.9	Inductive Reasoning with Bayesian Framework	41
3.9.1	Inductive Generalization	41
3.9.2	Taxonomy Learning	45
3.9.3	Feature Selection	46
3.10	Conclusion	47
4	Late Vision	51
4.1	Stereopsis and Depth Perception	51
4.2	Perception of Visual Quality	53
4.3	Perceptual Grouping	55
4.4	Foreground–Background Separation	59
4.5	Multi-stability	60
4.6	Object Recognition	61
4.6.1	In-Context Object Recognition	62
4.6.2	Synthesis of Bottom-Up and Top-Down Knowledge	64
4.6.3	Hierarchical Modeling	65
4.6.4	One-Shot Learning	66
4.7	Visual Aesthetics	67
4.8	Conclusion	69
5	Visual Attention	71
5.1	Modeling of Visual Attention	72
5.2	Models for Visual Attention	75
5.2.1	Cognitive Models	75
5.2.2	Information-Theoretic Models	77
5.2.3	Bayesian Models	78
5.2.4	Context-Based Models	79
5.2.5	Object-Based Models	81
5.3	Evaluation	82
5.4	Conclusion	84
6	Cognitive Architectures	87
6.1	Cognitive Modeling	88
6.1.1	Paradigms for Modeling Cognition	88
6.1.2	Levels of Abstraction	91
6.2	Desiderata for Cognitive Architectures	92
6.3	Memory Architecture	94

6.4	Taxonomies of Cognitive Architectures	97
6.5	Review of Cognitive Architectures	99
6.5.1	STAR: Selective Tuning Attentive Reference	100
6.5.2	LIDA: Learning Intelligent Distribution Agent	102
6.6	Biologically Inspired Cognitive Architectures	105
6.7	Conclusions	106
7	Knowledge Representation for Cognitive Vision	109
7.1	Classicist Approach to Knowledge Representation	109
7.1.1	First Order Logic	111
7.1.2	Semantic Networks	113
7.1.3	Frame-Based Representation	114
7.2	Symbol Grounding Problem	117
7.3	Perceptual Knowledge	118
7.3.1	Representing Perceptual Knowledge	119
7.3.2	Structural Description of Scenes	120
7.3.3	Qualitative Spatial and Temporal Relations	122
7.3.4	Inexact Spatiotemporal Relations	124
7.4	Unifying Conceptual and Perceptual Knowledge	127
7.5	Knowledge-Based Visual Data Processing	128
7.6	Conclusion	129
8	Deep Learning for Visual Cognition	131
8.1	A Brief Introduction to Deep Neural Networks	132
8.1.1	Fully Connected Networks	132
8.1.2	Convolutional Neural Networks	134
8.1.3	Recurrent Neural Networks	137
8.1.4	Siamese Networks	140
8.1.5	Graph Neural Networks	140
8.2	Modes of Learning with DNN	142
8.2.1	Supervised Learning	142
8.2.1.1	Image Segmentation	142
8.2.1.2	Object Detection	144
8.2.2	Unsupervised Learning with Generative Networks	144
8.2.3	Meta-Learning: Learning to Learn	146
8.2.3.1	Reinforcement Learning	148
8.2.3.2	One-Shot and Few-Shot Learning	148
8.2.3.3	Zero-Shot Learning	150
8.2.3.4	Incremental Learning	150
8.2.4	Multi-task Learning	152

8.3	Visual Attention	154
8.3.1	Recurrent Attention Models	155
8.3.2	Recurrent Attention Model for Video	158
8.4	Bayesian Inferencing with Neural Networks	159
8.5	Conclusion	160

9 Applications of Visual Cognition 163

9.1	Computational Photography	163
9.1.1	Color Enhancement	164
9.1.2	Intelligent Cropping	166
9.1.3	Face Beautification	167
9.2	Digital Heritage	168
9.2.1	Digital Restoration of Images	168
9.2.2	Curating Dance Archives	170
9.3	Social Robots	172
9.3.1	Dynamic and Shared Spaces	173
9.3.2	Recognition of Visual Cues	174
9.3.3	Attention to Socially Relevant Signals	175
9.4	Content Re-purposing	177
9.5	Conclusion	179

10 Conclusion 181

10.1	“What Is Cognitive Vision” Revisited	181
10.2	Divergence of Approaches	183
10.3	Convergence on the Anvil?	185

References 187

Index 215