

Contents

1	Is Human Vision Any Good?	1
	<i>Jan J. Koenderink</i>	
1.1	Introduction	1
1.1.1	What Is "Human Vision"?	1
1.1.2	How Is Human Vision Implemented?	2
1.2	Frameworks	5
1.2.1	The Spatial Framework	6
1.2.2	The Photometric Framework	9
1.3	A Case Study: "Shape from Shading"	13
1.3.1	The So Called "Shape from Shading Problem"	13
1.3.2	Setting up the Problem	15
1.3.3	The Local Shape from Shading Problem	18
1.4	Final Remarks	24
	References	25
2	Knowing a Good Feature When You See It: Ground Truth and Methodology to Evaluate Local Features for Recognition.....	27
	<i>Andrea Vedaldi, Haibin Ling, Stefano Soatto</i>	
2.1	Introduction	27
2.2	Empirical Studies of Local Features.....	29
2.2.1	Some Nomenclature	30
2.3	Constructing a Rigorous Ground Truth	30
2.3.1	Modeling the Detector	30
2.3.2	Modeling Correspondences	33
2.3.3	Ground-Truth Correspondences	34
2.3.4	Comparing Ground-Truth and Real-World Correspondences	36
2.3.5	The Data	37
2.4	Learning to Compare Invariant Features	40

2.4.1	Wide Baseline Motion Statistics	41
2.4.2	Learning to Rank Matches	41
2.4.3	Learning to Accept Matches	43
2.5	Discussion	44
	Appendix 1: Calculations	44
	Appendix 2: Algorithms	45
	References	48
3	Dynamic Graph Cuts and Their Applications in	
	Computer Vision	51
	<i>Pushmeet Kohli, Philip H.S. Torr</i>	
3.1	Introduction	51
3.1.1	Markov and Conditional Random Fields	52
3.2	Graph Cuts for Energy Minimization	54
3.2.1	The st-Mincut Problem	54
3.2.2	Formulating the Max-Flow Problem	55
3.2.3	Augmenting Paths, Residual Graphs	56
3.2.4	Minimizing Functions Using Graph Cuts	56
3.3	Minimizing Dynamic Energy Functions Using Dynamic	
	Graph Cuts	58
3.3.1	Dynamic Computation	58
3.3.2	Energy and Graph Reparameterization	60
3.4	Recycling Computation	61
3.4.1	Updating Residual Graphs	62
3.4.2	Computational Complexity of Update	
	Operations	64
3.5	Improving Performance by Recycling Search Trees	65
3.5.1	Reusing Search Trees	65
3.5.2	Tree Recycling for Dynamic Graph Cuts	66
3.6	Dynamic Image Segmentation	67
3.6.1	CRFs for Image Segmentation	68
3.6.2	Image Segmentation in Videos	70
3.6.3	Experimental Results	70
3.6.4	Reusing Flow vs. Reusing Search Trees	73
3.7	Simultaneous Segmentation and Pose Estimation of	
	Humans	74
3.7.1	Pose Specific CRF for Image Segmentation	78
3.7.2	Formulating the Pose Inference Problem	82
3.7.3	Experiments	85
3.7.4	Shape Priors for Reconstruction	87
3.7.5	Discussion	91
3.7.6	Summary and Future Work	92
3.8	Measuring Uncertainty in Graph Cut Solutions	93
3.8.1	Preliminaries	94
3.8.2	Computing Min-Marginals Using Graph Cuts	97

3.8.3	Min-Marginals and Flow Potentials.....	99
3.8.4	Extension to Multiple Labels	100
3.8.5	Minimizing Energy Function Projections Using Dynamic Graph Cuts	101
3.8.6	Computational Complexity and Experimental Evaluation	101
3.8.7	Applications of Min-Marginals	102
	References	104
4	Discriminative Graphical Models for Context-Based Classification	109
	<i>Sanjiv Kumar</i>	
4.1	Contextual Dependencies in Images.....	109
4.1.1	The Nature of Contextual Interactions	110
4.2	Markov Random Field (MRF)	112
4.3	Conditional Random Field (CRF)	113
4.3.1	Association Potential	115
4.3.2	Interaction Potential	117
4.4	Parameter Learning and Inference	119
4.4.1	Maximum Likelihood Parameter Learning.....	119
4.4.2	Inference	121
4.5	Extensions	121
4.5.1	Multiclass CRF	122
4.5.2	Hierarchical CRF	123
4.6	Applications	127
4.6.1	Man-Made Structure Detection	127
4.6.2	Image Classification and Contextual Object Detection	129
4.7	Related Work and Further Readings	133
	References	133
5	From the Subspace Methods to the Mutual Subspace Method.....	135
	<i>Ken-ichi Maeda</i>	
5.1	Introduction	135
5.2	The Subspace Methods	139
5.2.1	A Brief History of the Subspace Methods	139
5.2.2	Basic Idea	139
5.2.3	Subspace Construction	140
5.3	The Mutual Subspace Method	142
5.3.1	Basic Idea	142
5.3.2	Application to Chinese Character Recognition	144
5.3.3	Application to Face Recognition	147
5.3.4	Application to 3-D Face Recognition	148
5.4	Conclusion	154
	References	155

6	What, Where and Who? Telling the Story of an Image by Activity Classification, Scene Recognition and Object Categorization	157
	<i>Li Fei-Fei, Li-Jia Li</i>	
6.1	Introduction and Motivation	157
6.2	Overall Approach	159
6.3	Literature Review	160
6.4	The Integrative Model	160
6.4.1	Labeling an Unknown Image	163
6.5	Learning the Model	164
6.6	System Implementation	165
6.7	Experiments and Results	166
6.7.1	Dataset	166
6.7.2	Experimental Setup	166
6.7.3	Results	169
6.8	Conclusion	169
	References	170
7	Semantic Texton Forests	173
	<i>Matthew Johnson, Jamie Shotton</i>	
7.1	Introduction	173
7.2	Related Work	174
7.3	Randomized Decision Forests	175
7.3.1	Training the Forest	176
7.3.2	Experiments	180
7.4	Image Categorization	183
7.4.1	Tree Histograms and Pyramid Matching	184
7.4.2	Categorization Results	186
7.5	Semantic Segmentation	188
7.6	Soft Classification of Pixels	189
7.7	Image-Level Semantic Constraints	191
7.7.1	Categorization Results	192
7.7.2	The Image Level Prior	192
7.8	Compositional Constraints	194
7.9	Experiments	194
7.9.1	MSRC21 Dataset	196
7.9.2	VOC 2007 Segmentation Dataset	200
7.10	Discussion	200
	References	201
8	Multi-view Object Categorization and Pose Estimation	205
	<i>Silvio Savarese, Li Fei-Fei</i>	
8.1	Introduction	206
8.2	Literature Review	209
8.3	The Part-Based Multi-view Model	210

8.3.1	Overview	210
8.3.2	Canonical Parts and Linkage Structure	211
8.4	Building the Model	212
8.4.1	Extract Features	213
8.4.2	Form Parts	213
8.4.3	Find Canonical Parts Candidates	214
8.4.4	Create the Model	216
8.5	View Synthesis	217
8.5.1	Representing an Unseen View	217
8.6	Recognizing Object Class in Unseen Views	219
8.6.1	Extract Features and Get Part Candidates	220
8.6.2	Recognition Procedure: First Step	221
8.6.3	Recognition Procedure: Second Step	222
8.7	Experiments and Results	222
8.7.1	Experiment I: Comparison with Thomas et al. [63]	222
8.7.2	Experiment II: Detection and Pose Estimation Results on the Dataset in [54]	223
8.7.3	Experiment III: Detection and Pose Estimation Results on the Dataset in [55]	227
8.8	Conclusion	227
	References	228
9	A Vision-Based Remote Control	233
	Björn Stenger, Thomas Woodley, Roberto Cipolla	
9.1	Introduction	233
9.2	Prior Work	235
9.2.1	Hand Tracking for Human Computer Interfaces ...	235
9.2.2	Commercial Gesture Interface Systems	238
9.2.3	Visual Tracking of a Single Object	239
9.3	Tracking with Multiple Observers	240
9.3.1	Observation Models	240
9.3.2	Evaluating Single Observers	243
9.3.3	Evaluating Multiple Observers	245
9.3.4	Parallel Evaluation	246
9.3.5	Cascaded Evaluation	247
9.3.6	Dynamic Model Discussion	247
9.3.7	Experimental Results	248
9.3.8	Individual Observers	248
9.3.9	Observer Combinations	249
9.3.10	Tracker Evaluation on Selected Combinations	252
9.4	Gesture Interface System	253
9.4.1	Visual Attention Mechanism	253
9.4.2	Tracking Mechanism	254
9.4.3	Selection Mechanisms	255

9.5	Summary and Conclusion	257
	References	258
10	Multi-view Multi-object Detection and Tracking	263
	Murtaza Taj, Andrea Cavallaro	
10.1	Introduction	263
10.2	Problem Formulation	265
10.3	Calibration and Fusion	266
	10.3.1 Single-Level Homography	266
	10.3.2 Multi-Level Homography	268
10.4	Track-First Approaches	269
	10.4.1 Independent Tracking	271
	10.4.2 Collaborative Tracking	271
10.5	Fuse-First Approaches	272
	10.5.1 Detection-Based Tracking	273
	10.5.2 Track-Before-Detect	276
10.6	Conclusions	278
	References	278
11	Shape from Photographs: A Multi-view Stereo Pipeline	281
	Carlos Hernández, George Vogiatzis	
11.1	Introduction	281
11.2	Multi-view Stereo Pipeline: From Photographs to 3D Models	284
11.3	Computing Photo-Consistency from a Set of Calibrated Photographs	285
	11.3.1 Normalized Cross Correlation for Depth-Map Computation	287
	11.3.2 Depth Map Estimation	288
	11.3.3 Photo-Consistency 3D Map from a Set of Depth-Maps	292
11.4	Extracting a 3D Surface from a 3D Map of Photo-Consistency	292
	11.4.1 Multi-view Stereo Using Multi-resolution Graph-Cuts	292
	11.4.2 Discontinuity Cost from a Set of Depth-Maps	294
	11.4.3 Graph Structure	295
	11.4.4 Labeling Cost from a Set of Depth-Maps	295
	11.4.5 Probabilistic Fusion of Depth Sensors	296
	11.4.6 Deformable Models	299
11.5	Experiments	302
	11.5.1 Depth Map Evaluation	302
	11.5.2 Multi-view Stereo Evaluation	304

11.5.3	Digitizing Works of Art	304
11.6	Discussion	307
	References	309
12	Practical 3D Reconstruction Based on Photometric Stereo	313
	<i>George Vogiatzis, Carlos Hernández</i>	
12.1	Introduction	313
12.2	Photometric Stereo with Coloured Light	314
12.2.1	Classic Three-Source Photometric Stereo	315
12.2.2	Multi-spectral Sources and Sensors	316
12.2.3	Calibration	317
12.2.4	Comparison with Photometric Stereo	317
12.2.5	The Problem of Shadows	318
12.2.6	Facial Capture Experiments	325
12.2.7	Related Work	328
12.3	Multi-view Photometric Stereo	330
12.3.1	Related Work	331
12.3.2	Algorithm	333
12.3.3	Experiments	338
	References	342
	Index	347