

Contents

Preface	ix
1 Introduction	1
1.1 What is computer vision?	3
1.2 A brief history	9
1.3 Book overview	18
1.4 Sample syllabus	24
1.5 A note on notation	25
1.6 Additional reading	26
2 Image formation	27
2.1 Geometric primitives and transformations	29
2.1.1 2D transformations	32
2.1.2 3D transformations	35
2.1.3 3D rotations	36
2.1.4 3D to 2D projections	41
2.1.5 Lens distortions	51
2.2 Photometric image formation	53
2.2.1 Lighting	53
2.2.2 Reflectance and shading	54
2.2.3 Optics	59
2.3 The digital camera	63
2.3.1 Sampling and aliasing	67
2.3.2 Color	69
2.3.3 Compression	78
2.4 Additional reading	80
2.5 Exercises	80
3 Image processing	85
3.1 Point operators	87
3.1.1 Pixel transforms	89
3.1.2 Color transforms	90
3.1.3 Compositing and matting	91
3.1.4 Histogram equalization	92
3.1.5 <i>Application: Tonal adjustment</i>	95
3.2 Linear filtering	95
3.2.1 Separable filtering	99

3.2.2	Examples of linear filtering	100
3.2.3	Band-pass and steerable filters	101
3.3	More neighborhood operators	105
3.3.1	Non-linear filtering	105
3.3.2	Bilateral filtering	107
3.3.3	Binary image processing	110
3.4	Fourier transforms	113
3.4.1	Two-dimensional Fourier transforms	115
3.4.2	<i>Application:</i> Sharpening, blur, and noise removal	118
3.5	Pyramids and wavelets	119
3.5.1	Interpolation	119
3.5.2	Decimation	122
3.5.3	Multi-resolution representations	123
3.5.4	Wavelets	128
3.5.5	<i>Application:</i> Image blending	132
3.6	Geometric transformations	135
3.6.1	Parametric transformations	135
3.6.2	Mesh-based warping	140
3.6.3	<i>Application:</i> Feature-based morphing	142
3.7	Additional reading	143
3.8	Exercises	144
4	Model fitting and optimization	153
4.1	Scattered data interpolation	155
4.1.1	Radial basis functions	157
4.1.2	Overfitting and underfitting	159
4.1.3	Robust data fitting	162
4.2	Variational methods and regularization	163
4.2.1	Discrete energy minimization	166
4.2.2	Total variation	168
4.2.3	Bilateral solver	168
4.2.4	<i>Application:</i> Interactive colorization	169
4.3	Markov random fields	170
4.3.1	Conditional random fields	177
4.3.2	<i>Application:</i> Interactive segmentation	181
4.4	Additional reading	184
4.5	Exercises	185
5	Deep Learning	187
5.1	Supervised learning	191
5.1.1	Nearest neighbors	192
5.1.2	Bayesian classification	194
5.1.3	Logistic regression	198
5.1.4	Support vector machines	199
5.1.5	Decision trees and forests	202
5.2	Unsupervised learning	205
5.2.1	Clustering	205
5.2.2	K-means and Gaussians mixture models	206

5.2.3	Principal component analysis	209
5.2.4	Manifold learning	211
5.2.5	Semi-supervised learning	212
5.3	Deep neural networks	214
5.3.1	Weights and layers	215
5.3.2	Activation functions	217
5.3.3	Regularization and normalization	219
5.3.4	Loss functions	223
5.3.5	Backpropagation	226
5.3.6	Training and optimization	228
5.4	Convolutional neural networks	231
5.4.1	Pooling and unpooling	234
5.4.2	<i>Application: Digit classification</i>	237
5.4.3	Network architectures	238
5.4.4	Model zoos	242
5.4.5	Visualizing weights and activations	244
5.4.6	Adversarial examples	248
5.4.7	Self-supervised learning	249
5.5	More complex models	252
5.5.1	Three-dimensional CNNs	252
5.5.2	Recurrent neural networks	255
5.5.3	Transformers	257
5.5.4	Generative models	261
5.6	Additional reading	267
5.7	Exercises	268
6	Recognition	273
6.1	Instance recognition	276
6.2	Image classification	278
6.2.1	Feature-based methods	278
6.2.2	Deep networks	285
6.2.3	<i>Application: Visual similarity search</i>	287
6.2.4	Face recognition	289
6.3	Object detection	295
6.3.1	Face detection	295
6.3.2	Pedestrian detection	299
6.3.3	General object detection	301
6.4	Semantic segmentation	307
6.4.1	<i>Application: Medical image segmentation</i>	310
6.4.2	Instance segmentation	311
6.4.3	Panoptic segmentation	312
6.4.4	<i>Application: Intelligent photo editing</i>	314
6.4.5	Pose estimation	315
6.5	Video understanding	316
6.6	Vision and language	319
6.7	Additional reading	326
6.8	Exercises	329

7	Feature detection and matching	333
7.1	Points and patches	335
7.1.1	Feature detectors	337
7.1.2	Feature descriptors	347
7.1.3	Feature matching	352
7.1.4	Large-scale matching and retrieval	358
7.1.5	Feature tracking	361
7.1.6	<i>Application: Performance-driven animation</i>	363
7.2	Edges and contours	364
7.2.1	Edge detection	364
7.2.2	Contour detection	368
7.2.3	<i>Application: Edge editing and enhancement</i>	372
7.3	Contour tracking	373
7.3.1	Snakes and scissors	373
7.3.2	Level Sets	379
7.3.3	<i>Application: Contour tracking and rotoscoping</i>	380
7.4	Lines and vanishing points	381
7.4.1	Successive approximation	381
7.4.2	Hough transforms	381
7.4.3	Vanishing points	384
7.5	Segmentation	386
7.5.1	Graph-based segmentation	388
7.5.2	Mean shift	389
7.5.3	Normalized cuts	391
7.6	Additional reading	393
7.7	Exercises	395
8	Image alignment and stitching	401
8.1	Pairwise alignment	403
8.1.1	2D alignment using least squares	403
8.1.2	<i>Application: Panography</i>	405
8.1.3	Iterative algorithms	406
8.1.4	Robust least squares and RANSAC	408
8.1.5	3D alignment	410
8.2	Image stitching	411
8.2.1	Parametric motion models	412
8.2.2	<i>Application: Whiteboard and document scanning</i>	414
8.2.3	Rotational panoramas	414
8.2.4	Gap closing	416
8.2.5	<i>Application: Video summarization and compression</i>	417
8.2.6	Cylindrical and spherical coordinates	418
8.3	Global alignment	421
8.3.1	Bundle adjustment	421
8.3.2	Parallax removal	424
8.3.3	Recognizing panoramas	425
8.4	Compositing	426
8.4.1	Choosing a compositing surface	426
8.4.2	Pixel selection and weighting (deghosting)	430

8.4.3	<i>Application: Photomontage</i>	435
8.4.4	Blending	435
8.5	Additional reading	437
8.6	Exercises	438
9	Motion estimation	443
9.1	Translational alignment	445
9.1.1	Hierarchical motion estimation	448
9.1.2	Fourier-based alignment	449
9.1.3	Incremental refinement	451
9.2	Parametric motion	455
9.2.1	<i>Application: Video stabilization</i>	457
9.2.2	Spline-based motion	459
9.2.3	<i>Application: Medical image registration</i>	461
9.3	Optical flow	461
9.3.1	Deep learning approaches	466
9.3.2	<i>Application: Rolling shutter wobble removal</i>	468
9.3.3	Multi-frame motion estimation	468
9.3.4	<i>Application: Video denoising</i>	469
9.4	Layered motion	470
9.4.1	<i>Application: Frame interpolation</i>	473
9.4.2	Transparent layers and reflections	474
9.4.3	Video object segmentation	476
9.4.4	Video object tracking	477
9.5	Additional reading	478
9.6	Exercises	479
10	Computational photography	483
10.1	Photometric calibration	486
10.1.1	Radiometric response function	486
10.1.2	Noise level estimation	488
10.1.3	Vignetting	489
10.1.4	Optical blur (spatial response) estimation	491
10.2	High dynamic range imaging	494
10.2.1	Tone mapping	499
10.2.2	<i>Application: Flash photography</i>	506
10.3	Super-resolution, denoising, and blur removal	508
10.3.1	Color image demosaicing	515
10.3.2	Lens blur (bokeh)	517
10.4	Image matting and compositing	518
10.4.1	Blue screen matting	519
10.4.2	Natural image matting	521
10.4.3	Optimization-based matting	524
10.4.4	Smoke, shadow, and flash matting	527
10.4.5	Video matting	528
10.5	Texture analysis and synthesis	529
10.5.1	<i>Application: Hole filling and inpainting</i>	531
10.5.2	<i>Application: Non-photorealistic rendering</i>	532

10.5.3	Neural style transfer and semantic image synthesis	534
10.6	Additional reading	537
10.7	Exercises	538
11	Structure from motion and SLAM	543
11.1	Geometric intrinsic calibration	545
11.1.1	Vanishing points	547
11.1.2	<i>Application</i> : Single view metrology	548
11.1.3	Rotational motion	549
11.1.4	Radial distortion	550
11.2	Pose estimation	552
11.2.1	Linear algorithms	552
11.2.2	Iterative non-linear algorithms	554
11.2.3	<i>Application</i> : Location recognition	555
11.2.4	Triangulation	558
11.3	Two-frame structure from motion	560
11.3.1	Eight, seven, and five-point algorithms	560
11.3.2	Special motions and structures	564
11.3.3	Projective (uncalibrated) reconstruction	565
11.3.4	Self-calibration	566
11.3.5	<i>Application</i> : View morphing	568
11.4	Multi-frame structure from motion	568
11.4.1	Factorization	568
11.4.2	Bundle adjustment	570
11.4.3	Exploiting sparsity	571
11.4.4	<i>Application</i> : Match move	574
11.4.5	Uncertainty and ambiguities	575
11.4.6	<i>Application</i> : Reconstruction from internet photos	576
11.4.7	Global structure from motion	578
11.4.8	Constrained structure and motion	580
11.5	Simultaneous localization and mapping (SLAM)	583
11.5.1	<i>Application</i> : Autonomous navigation	585
11.5.2	<i>Application</i> : Smartphone augmented reality	587
11.6	Additional reading	588
11.7	Exercises	590
12	Depth estimation	595
12.1	Epipolar geometry	599
12.1.1	Rectification	600
12.1.2	Plane sweep	602
12.2	Sparse correspondence	604
12.2.1	3D curves and profiles	604
12.3	Dense correspondence	606
12.3.1	Similarity measures	607
12.4	Local methods	609
12.4.1	Sub-pixel estimation and uncertainty	610
12.4.2	<i>Application</i> : Stereo-based head tracking	611
12.5	Global optimization	612

12.5.1	Dynamic programming	614
12.5.2	Segmentation-based techniques	616
12.5.3	<i>Application: Z-keying and background replacement</i>	617
12.6	Deep neural networks	618
12.7	Multi-view stereo	620
12.7.1	Scene flow	624
12.7.2	Volumetric and 3D surface reconstruction	624
12.7.3	Shape from silhouettes	630
12.8	Monocular depth estimation	632
12.9	Additional reading	634
12.10	Exercises	635
13	3D reconstruction	639
13.1	Shape from X	641
13.1.1	Shape from shading and photometric stereo	642
13.1.2	Shape from texture	645
13.1.3	Shape from focus	646
13.2	3D scanning	647
13.2.1	Range data merging	650
13.2.2	<i>Application: Digital heritage</i>	654
13.3	Surface representations	654
13.3.1	Surface interpolation	655
13.3.2	Surface simplification	656
13.3.3	Geometry images	656
13.4	Point-based representations	657
13.5	Volumetric representations	658
13.5.1	Implicit surfaces and level sets	658
13.6	Model-based reconstruction	660
13.6.1	Architecture	660
13.6.2	Facial modeling and tracking	663
13.6.3	<i>Application: Facial animation</i>	665
13.6.4	Human body modeling and tracking	668
13.7	Recovering texture maps and albedos	674
13.7.1	Estimating BRDFs	675
13.7.2	<i>Application: 3D model capture</i>	676
13.8	Additional reading	677
13.9	Exercises	679
14	Image-based rendering	681
14.1	View interpolation	683
14.1.1	View-dependent texture maps	685
14.1.2	<i>Application: Photo Tourism</i>	686
14.2	Layered depth images	688
14.2.1	Impostors, sprites, and layers	688
14.2.2	<i>Application: 3D photography</i>	690
14.3	Light fields and Lumigraphs	693
14.3.1	Unstructured Lumigraph	696
14.3.2	Surface light fields	696

14.3.3	<i>Application: Concentric mosaics</i>	698
14.3.4	<i>Application: Synthetic re-focusing</i>	698
14.4	Environment mattes	699
14.4.1	Higher-dimensional light fields	700
14.4.2	The modeling to rendering continuum	701
14.5	Video-based rendering	701
14.5.1	Video-based animation	702
14.5.2	Video textures	703
14.5.3	<i>Application: Animating pictures</i>	705
14.5.4	3D and free-viewpoint Video	706
14.5.5	<i>Application: Video-based walkthroughs</i>	708
14.6	Neural rendering	711
14.7	Additional reading	718
14.8	Exercises	719
15	Conclusion	723
A	Linear algebra and numerical techniques	727
A.1	Matrix decompositions	728
A.1.1	Singular value decomposition	728
A.1.2	Eigenvalue decomposition	729
A.1.3	QR factorization	731
A.1.4	Cholesky factorization	732
A.2	Linear least squares	733
A.2.1	Total least squares	734
A.3	Non-linear least squares	736
A.4	Direct sparse matrix techniques	737
A.4.1	Variable reordering	737
A.5	Iterative techniques	738
A.5.1	Conjugate gradient	739
A.5.2	Preconditioning	740
A.5.3	Multigrid	741
B	Bayesian modeling and inference	743
B.1	Estimation theory	745
B.2	Maximum likelihood estimation and least squares	747
B.3	Robust statistics	748
B.4	Prior models and Bayesian inference	750
B.5	Markov random fields	751
B.6	Uncertainty estimation (error analysis)	752
C	Supplementary material	755
C.1	Datasets and benchmarks	756
C.2	Software	761
C.3	Slides and lectures	768
	References	769
	Index	905