

Computer Vision Metrics

Survey, Taxonomy, and Analysis



Scott Krig



Apress
open

Computer Vision Metrics: Survey, Taxonomy, and Analysis

Scott Krig

Copyright © 2014 by Apress Media, LLC, all rights reserved

ApressOpen Rights: You have the right to copy, use and distribute this Work in its entirety, electronically without modification, for non-commercial purposes only. However, you have the additional right to use or alter any source code in this Work for any commercial or non-commercial purpose which must be accompanied by the licenses in (2) and (3) below to distribute the source code for instances of greater than 5 lines of code. Licenses (1), (2) and (3) below and the intervening text must be provided in any use of the text of the Work and fully describes the license granted herein to the Work.

(1) **License for Distribution of the Work:** This Work is copyrighted by Apress Media, LLC, all rights reserved. Use of this Work other than as provided for in this license is prohibited. By exercising any of the rights herein, you are accepting the terms of this license. You have the non-exclusive right to copy, use and distribute this English language Work in its entirety, electronically without modification except for those modifications necessary for formatting on specific devices, for all non-commercial purposes, in all media and formats known now or hereafter. While the advice and information in this Work are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

If your distribution is solely Apress source code or uses Apress source code intact, the following licenses (2) and (3) must accompany the source code. If your use is an adaptation of the source code provided by Apress in this Work, then you must use only license (3).

(2) **License for Direct Reproduction of Apress Source Code:** This source code, from *Computer Vision Metrics: Survey, Taxonomy, and Analysis, ISBN 978-1-4302-5929-9* is copyrighted by Apress Media, LLC, all rights reserved. Any direct reproduction of this Apress source code is permitted but must contain this license. The following license must be provided for any use of the source code from this product of greater than 5 lines wherein the code is adapted or altered from its original Apress form. This Apress code is presented AS IS and Apress makes no claims to, representations or warranties as to the function, usability, accuracy or usefulness of this code.

(3) **License for Distribution of Adaptation of Apress Source Code:** Portions of the source code provided are used or adapted from *Computer Vision Metrics: Survey, Taxonomy, and Analysis, ISBN 978-1-4302-5929-9* copyright Apress Media LLC. Any use or reuse of this Apress source code must contain this License. This Apress code is made available at Apress.com/978143026136-0 as is and Apress makes no claims to, representations or warranties as to the function, usability, accuracy or usefulness of this code.

ISBN-13 (pbk): 978-1-4302-5929-9

ISBN-13 (electronic): 978-1-4302-5930-5

Trademarked names, logos, and images may appear in this book. Rather than use a trademark symbol with every occurrence of a trademarked name, logo, or image we use the names, logos, and images only in an editorial fashion and to the benefit of the trademark owner, with no intention of infringement of the trademark.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

President and Publisher: Paul Manning
Lead Editors: Steve Weiss (Apress); Stuart Douglas (Intel)
Technical Reviewers: Sanjay Addicam and Shahzad Malik
Coordinating Editor: Melissa Maldonado
Cover Designer: Anna Ishchenko

Distributed to the book trade worldwide by Springer Science+Business Media New York, 233 Spring Street, 6th Floor, New York, NY 10013. Phone 1-800-SPRINGER, fax (201) 348-4505, e-mail orders-ny@springer-sbm.com, or visit www.springeronline.com.

For information on translations, please e-mail rights@apress.com, or visit www.apress.com.

About ApressOpen

What Is ApressOpen?

- ApressOpen is an open-access book program that publishes high-quality technical and business information.
- ApressOpen eBooks are available for global, free, noncommercial use.
- ApressOpen eBooks are available in PDF, ePub, and Mobi formats.
- The user friendly ApressOpen free eBook license is presented on the copyright page of this book.

Thanks to my wife Janie, family, and parents for being part of my life.

Contents at a Glance

About the Author	xxvii
Acknowledgments	xxix
Introduction	xxxix
■ Chapter 1: Image Capture and Representation.....	1
■ Chapter 2: Image Pre-Processing.....	39
■ Chapter 3: Global and Regional Features	85
■ Chapter 4: Local Feature Design Concepts, Classification, and Learning.....	131
■ Chapter 5: Taxonomy of Feature Description Attributes.....	191
■ Chapter 6: Interest Point Detector and Feature Descriptor Survey.....	217
■ Chapter 7: Ground Truth Data, Content, Metrics, and Analysis... 	283
■ Chapter 8: Vision Pipelines and Optimizations.....	313
■ Appendix A: Synthetic Feature Analysis.....	365
■ Appendix B: Survey of Ground Truth Datasets.....	401
■ Appendix C: Imaging and Computer Vision Resources.....	411
■ Appendix D: Extended SDM Metrics	419
■ Bibliography	437
Index.....	465

Contents

- About the Authorxxvii**
- Acknowledgmentsxxix**
- Introductionxxxi**
- Chapter 1: Image Capture and Representation..... 1**
 - Image Sensor Technology 1**
 - Sensor Materials 2
 - Sensor Photo-Diode Cells 3
 - Sensor Configurations: Mosaic, Foveon, BSI 4
 - Dynamic Range and Noise 6
 - Sensor Processing..... 6
 - De-Mosaicking 6
 - Dead Pixel Correction 7
 - Color and Lighting Corrections 7
 - Geometric Corrections..... 7
 - Cameras and Computational Imaging 8**
 - Overview of Computational Imaging 8
 - Single-Pixel Computational Cameras 9
 - 2D Computational Cameras..... 10
 - 3D Depth Camera Systems..... 12
 - Binocular Stereo 14
 - Structured and Coded Light 17
 - Optical Coding: Diffraction Gratings 19

Time-of-Flight Sensors	20
Array Cameras	22
Radial Cameras	22
Plenoptics: Light Field Cameras	23
3D Depth Processing	24
Overview of Methods.....	25
Problems in Depth Sensing and Processing.....	25
The Geometric Field and Distortions	26
The Horopter Region, Panum’s Area, and Depth Fusion.....	26
Cartesian vs. Polar Coordinates: Spherical Projective Geometry.....	27
Depth Granularity.....	28
Correspondence	29
Holes and Occlusion.....	30
Surface Reconstruction and Fusion	30
Noise.....	32
Monocular Depth Processing	32
Multi-View Stereo	32
Sparse Methods: PTAM.....	33
Dense Methods: DTAM	34
Optical Flow, SLAM, and SFM	34
3D Representations: Voxels, Depth Maps, Meshes, and Point Clouds ...	35
Summary.....	37
■ Chapter 2: Image Pre-Processing.....	39
Perspectives on Image Processing	39
Problems to Solve During Image Pre-Processing	40
Vision Pipelines and Image Pre-Processing	40
Corrections	42
Enhancements.....	43

Preparing Images for Feature Extraction	43
Local Binary Family Pre-Processing.....	43
Spectra Family Pre-Processing.....	45
Basis Space Family Pre-Processing.....	46
Polygon Shape Family Pre-Processing.....	47
The Taxonomy of Image Processing Methods	50
Point.....	50
Line.....	50
Area.....	51
Algorithmic.....	51
Data Conversions.....	51
Colorimetry	51
Overview of Color Management Systems.....	52
Illuminants, White Point, Black Point, and Neutral Axis.....	53
Device Color Models.....	54
Color Spaces and Color Perception.....	55
Gamut Mapping and Rendering Intent.....	55
Practical Considerations for Color Enhancements.....	56
Color Accuracy and Precision.....	57
Spatial Filtering	57
Convolutional Filtering and Detection.....	58
Kernel Filtering and Shape Selection.....	60
Shape Selection or Forming Kernels.....	61
Point Filtering.....	61
Noise and Artifact Filtering.....	63
Integral Images and Box Filters.....	63
Edge Detectors	64
Kernel Sets: Sobel, Scharr, Prewitt, Roberts, Kirsch, Robinson, and Frei-Chen.....	64
Canny Detector.....	66

Transform Filtering, Fourier, and Others..... 67

- Fourier Transform Family 67
 - Fundamentals 67
 - Fourier Family of Transforms..... 70
 - Other Transforms..... 70

Morphology and Segmentation 71

- Binary Morphology 72
- Gray Scale and Color Morphology 73
- Morphology Optimizations and Refinements..... 73
- Euclidean Distance Maps 74
- Super-Pixel Segmentation 74
 - Graph-based Super-Pixel Methods 75
 - Gradient-Ascent-Based Super-Pixel Methods..... 75
- Depth Segmentation..... 76
- Color Segmentation 77

Thresholding 77

- Global Thresholding 77
 - Histogram Peaks and Valleys, and Hysteresis Thresholds..... 78
 - LUT Transforms, Contrast Remapping 78
 - Histogram Equalization and Specification 79
 - Global Auto Thresholding 80
- Local Thresholding 81
 - Local Histogram Equalization..... 81
 - Integral Image Contrast Filters..... 81
 - Local Auto Threshold Methods..... 82

Summary..... 83

■ Chapter 3: Global and Regional Features 85

Historical Survey of Features 85

 Key Ideas: Global, Regional, and Local 86

 1960s, 1970s, 1980s—Whole-Object Approaches 87

 Early 1990s—Partial-Object Approaches 87

 Mid-1990s—Local Feature Approaches 87

 Late 1990s—Classified Invariant Local Feature Approaches 88

 Early 2000s—Scene and Object Modeling Approaches 88

 Mid-2000s—Finer-Grain Feature and Metric Composition Approaches 88

 Post-2010—Multi-Modal Feature Metrics Fusion 88

 Textural Analysis 89

 1950s thru 1970s—Global Uniform Texture Metrics 90

 1980s—Structural and Model-Based Approaches for Texture Classification 91

 1990s—Optimizations and Refinements to Texture Metrics 91

 2000 to Today—More Robust Invariant Texture Metrics and 3D Texture 92

 Statistical Methods 92

Texture Region Metrics 93

 Edge Metrics 93

 Edge Density 94

 Edge Contrast 94

 Edge Entropy 94

 Edge Directivity 95

 Edge Linearity 95

 Edge Periodicity 95

 Edge Size 95

 Edge Primitive Length Total 96

 Cross-Correlation and Auto-Correlation 96

 Fourier Spectrum, Wavelets, and Basis Signatures 96

Co-Occurrence Matrix, Haralick Features.....	97
Extended SDM Metrics	100
Metric 1: Centroid.....	101
Metric 2: Total Coverage	101
Metric 3: Low-Frequency Coverage	102
Metric 4: Corrected Coverage	102
Metric 5: Total Power.....	102
Metric 6: Relative Power.....	103
Metric 7: Locus Mean Density.....	103
Metric 8: Locus Length	103
Metric 9: Bin Mean Density.....	104
Metric 10: Containment	104
Metric 11. Linearity	104
Metric 12: Linearity Strength	106
Laws Texture Metrics	106
LBP Local Binary Patterns	108
Dynamic Textures	108
Statistical Region Metrics	109
Image Moment Features	109
Point Metric Features	110
Global Histograms	112
Local Region Histograms.....	113
Scatter Diagrams, 3D Histograms	113
Multi-Resolution, Multi-Scale Histograms.....	117
Radial Histograms	118
Contour or Edge Histograms	118
Basis Space Metrics.....	118
Fourier Description.....	121
Walsh–Hadamard Transform	122

HAAR Transform.....	123
Slant Transform	123
Zernike Polynomials	124
Steerable Filters	124
Karhunen-Loeve Transform and Hotelling Transform	125
Wavelet Transform and Gabor Filters	125
Gabor Functions	127
Hough Transform and Radon Transform	127
Summary.....	129
■ Chapter 4: Local Feature Design Concepts, Classification, and Learning.....	131
Local Features.....	132
Detectors, Interest Points, Keypoints, Anchor Points, Landmarks	132
Descriptors, Feature Description, Feature Extraction.....	133
Sparse Local Pattern Methods	133
Local Feature Attributes	134
Choosing Feature Descriptors and Interest Points	134
Feature Descriptors and Feature Matching.....	134
Criteria for Goodness.....	134
Repeatability, Easy vs. Hard to Find	136
Distinctive vs. Indistinctive.....	137
Relative and Absolute Position	137
Matching Cost and Correspondence.....	137
Distance Functions.....	138
Early Work on Distance Functions	138
Euclidean or Cartesian Distance Metrics.....	139
Euclidean Distance	139
Squared Euclidean Distance	140

Cosine Distance or Similarity	140
Sum of Absolute Differences (SAD) or L1 Norm	140
Sum of Squared Differences (SSD) or L2 Norm	140
Correlation Distance	141
Hellinger Distance	141
Grid Distance Metrics	141
Manhattan Distance	141
Chebyshev Distance	142
Statistical Difference Metrics	142
Earth Movers Distance (EMD) or Wasserstein Metric	142
Mahalanobis Distance	143
Bray Curtis Distance	143
Canberra Distance	143
Binary or Boolean Distance Metrics	143
L0 Norm	143
Hamming Distance	144
Jaccard Similarity and Dissimilarity	144
Descriptor Representation	144
Coordinate Spaces, Complex Spaces	144
Cartesian Coordinates	145
Polar and Log Polar Coordinates	145
Radial Coordinates	145
Spherical Coordinates	146
Gauge Coordinates	146
Multivariate Spaces, Multimodal Data	146
Feature Pyramids	147
Descriptor Density	147
Interest Point and Descriptor Culling	147
Dense vs. Sparse Feature Description	148

Descriptor Shape Topologies..... 149

- Correlation Templates..... 149
- Patches and Shape..... 149
 - Single Patches, Sub-Patches 149
 - Deformable Patches 149
 - Multi-Patch Sets..... 150
 - TPLBP, FPLBP.....150
 - Strip and Radial Fan Shapes 151
 - D-NETS Strip Patterns.....151
- Object Polygon Shapes..... 152
 - Morphological Boundary Shapes 152
 - Texture Structure Shapes 153
 - Super-Pixel Similarity Shapes..... 153

Local Binary Descriptor Point-Pair Patterns..... 153

- FREAK Retinal Patterns 154
- Brisk Patterns..... 155
- ORB and BRIEF Patterns..... 156

Descriptor Discrimination 157

- Spectra Discrimination 158
- Region, Shapes, and Pattern Discrimination 159
- Geometric Discrimination Factors 160
- Feature Visualization to Evaluate Discrimination 160
 - Discrimination via Image Reconstruction from HOG 160
 - Discrimination via Image Reconstruction from Local Binary Patterns 161
 - Discrimination via Image Reconstruction from SIFT Features 162
- Accuracy, Trackability 163
- Accuracy Optimizations, Sub-Region Overlap, Gaussian Weighting, and Pooling..... 165
- Sub-Pixel Accuracy..... 165

Search Strategies and Optimizations	166
Dense Search	166
Grid Search.....	166
Multi-Scale Pyramid Search.....	167
Scale Space and Image Pyramids	168
Feature Pyramids	169
Sparse Predictive Search and Tracking	170
Tracking Region-Limited Search	170
Segmentation Limited Search	171
Depth or Z Limited Search.....	171
Computer Vision, Models, Organization.....	172
Feature Space	172
Object Models.....	173
Constraints	175
Selection of Detectors and Features	175
Manually Designed Feature Detectors.....	175
Statistically Designed Feature Detectors.....	175
Learned Features	176
Overview of Training.....	176
Classification of Features and Objects	177
Group Distance: Clustering, Training, and Statistical Learning	177
Group Distance: Clustering Methods Survey, KNN, RANSAC, K-Means, GMM, SVM, Others.....	178
Classification Frameworks, REIN, MOPED	180
Kernel Machines.....	181
Boosting, Weighting.....	181
Selected Examples of Classification.....	182

Feature Learning, Sparse Coding, Convolutional Networks.....	183
Terminology: Codebooks, Visual Vocabulary, Bag of Words, Bag of Features	183
Sparse Coding	184
Visual Vocabularies.....	185
Learned Detectors via Convolutional Filter Masks	186
Convolutional Neural Networks, Neural Networks	186
Deep Learning, Pooling, Trainable Feature Hierarchies	188
Summary.....	188
■ Chapter 5: Taxonomy of Feature Description Attributes.....	191
Feature Descriptor Families	192
Prior Work on Computer Vision Taxonomies	193
Robustness and Accuracy	194
General Robustness Taxonomy	195
Illumination.....	196
Color Criteria	196
Incompleteness	197
Resolution and Accuracy	197
Geometric Distortion.....	198
Efficiency Variables, Costs and Benefits.....	199
Discrimination and Uniqueness.....	199
General Vision Metrics Taxonomy.....	199
Feature Descriptor Family	201
Spectra Dimensions	201
Spectra Type.....	201
Interest Point	205
Storage Formats.....	206
Data Types	206

Descriptor Memory.....	207
Feature Shapes	207
Feature Pattern.....	207
Feature Density	208
Feature Search Methods	209
Pattern Pair Sampling	210
Pattern Region Size	211
Distance Function.....	211
Euclidean or Cartesian Distance Family	211
Grid Distance Family.....	212
Statistical Distance Family.....	212
Binary or Boolean Distance Family.....	212
Feature Metric Evaluation	212
Efficiency Variables, Costs and Benefits.....	213
Image Reconstruction Efficiency Metric.....	213
Example Feature Metric Evaluations	213
SIFT Example	213
VISION METRIC TAXONOMY FME	214
GENERAL ROBUSTNESS ATTRIBUTES	214
LBP Example	214
VISION METRIC TAXONOMY FME	214
GENERAL ROBUSTNESS ATTRIBUTES	215
Shape Factors Example	215
VISION METRIC TAXONOMY FME	215
GENERAL ROBUSTNESS ATTRIBUTES	216
Summary.....	216

■ Chapter 6: Interest Point Detector and Feature Descriptor Survey..... 217

Interest Point Tuning 218

Interest Point Concepts 218

Interest Point Method Survey..... 221

 Laplacian and Laplacian of Gaussian 222

 Moravac Corner Detector 222

 Harris Methods, Harris-Stephens, Shi-Tomasi, and Hessian-Type Detectors 222

 Hessian Matrix Detector and Hessian-Laplace..... 223

 Difference of Gaussians 223

 Salient Regions..... 224

 SUSAN, and Trajkovic and Hedly..... 224

 Fast, Faster, AGHAST 225

 Local Curvature Methods 226

 Morphological Interest Regions..... 227

Feature Descriptor Survey 227

 Local Binary Descriptors 228

 Local Binary Patterns..... 228

 Neighborhood Comparison..... 231

 Histogram Composition..... 231

 Optionally Normalization 232

 Descriptor Concatenation..... 232

 Rotation Invariant LBP (RILBP) 232

 Dynamic Texture Metric Using 3D LBPs 233

 Volume LBP (VLBP) 233

 LPB-TOP 234

 Other LBP Variants..... 234

Census.....	237
Modified Census Transform.....	237
BRIEF.....	238
ORB.....	238
BRISK.....	239
FREAK.....	240
Spectra Descriptors	241
SIFT	241
Create a Scale Space Pyramid.....	242
Identify Scale-Invariant Interest Points.....	244
Create Feature Descriptors.....	244
SIFT-PCA.....	246
SIFT-GLOH.....	246
SIFT-SIFER Retrofit.....	247
SIFT CS-LBP Retrofit.....	247
RootSIFT Retrofit.....	248
CenSurE and STAR.....	249
Correlation Templates.....	251
HAAR Features.....	252
Viola Jones with HAAR-Like Features	254
SURF.....	254
Variations on SURF.....	256
Histogram of Gradients (HOG) and Variants.....	257
PHOG and Related Methods.....	258
Daisy and O-Daisy	260
CARD	261
Robust Fast Feature Matching	263
RIFF, CHOG.....	264
Chain Code Histograms	266

D-NETS	266
Local Gradient Pattern.....	267
Local Phase Quantization	268
Basis Space Descriptors	269
Fourier Descriptors.....	269
Other Basis Functions for Descriptor Building	271
Sparse Coding Methods	271
Examples of Sparse Coding Methods.....	271
Polygon Shape Descriptors	272
MSER Method.....	273
Object Shape Metrics for Blobs and Polygons.....	274
Shape Context	277
3D, 4D, Volumetric, and Multimodal Descriptors	278
3D HOG	279
HON 4D.....	280
3D SIFT	280
Summary.....	282
■ Chapter 7: Ground Truth Data, Content, Metrics, and Analysis... 283	
What Is Ground Truth Data?	284
Previous Work on Ground Truth Data: Art vs. Science	286
General Measures of Quality Performance.....	286
Measures of Algorithm Performance.....	286
Rosin’s Work on Corners	287
Key Questions For Constructing Ground Truth Data	289
Content: Adopt, Modify, or Create	289
Survey Of Available Ground Truth Data.....	289
Fitting Data to Algorithms.....	290

Scene Composition and Labeling	291
Composition	292
Labeling	293
Defining the Goals and Expectations.....	294
Mikolajczyk and Schmid Methodology	295
Open Rating Systems	295
Corner Cases and Limits	295
Interest Points and Features	295
Robustness Criteria for Ground Truth Data.....	296
Illustrated Robustness Criteria	296
Using Robustness Criteria for Real Applications	299
Pairing Metrics with Ground Truth.....	300
Pairing and Tuning Interest Points, Features, and Ground Truth.....	301
Examples Using The General Vision Taxonomy.....	301
Synthetic Feature Alphabets	303
Goals for the Synthetic Dataset.....	304
Accuracy of Feature Detection via Location Grid	305
Rotational Invariance via Rotated Image Set	305
Scale Invariance via Thickness and Bounding Box Size	305
Noise and Blur Invariance	305
Repeatability	306
Real Image Overlays of Synthetic Features	306
Synthetic Interest Point Alphabet	306
Synthetic Corner Alphabet	307
Hybrid Synthetic Overlays on Real Images.....	309
Method for Creating the Overlays.....	310
Summary.....	310

■ Chapter 8: Vision Pipelines and Optimizations..... 313

Stages, Operations, and Resources 314

Compute Resource Budgets..... 315

 Compute Units, ALUs, and Accelerators..... 317

 Power Use 318

 Memory Use 319

 I/O Performance 322

The Vision Pipeline Examples..... 323

Automobile Recognition 323

 Segmenting the Automobiles 325

 Matching the Paint Color 326

 Measuring the Automobile Size and Shape 326

 Feature Descriptors 327

 Calibration, Set-up, and Ground Truth Data 328

 Pipeline Stages and Operations 329

 Operations and Compute Resources 330

 Criteria for Resource Assignments 330

Face, Emotion, and Age Recognition 331

 Calibration and Ground Truth Data 333

 Interest Point Position Prediction..... 334

 Segmenting the Head and Face Using the Bounding Box 335

 Face Landmark Identification and Compute Features..... 336

 Pipeline Stages and Operations 338

 Operations and Compute Resources 339

 Criteria for Resource Assignments 339

Image Classification 340

 Segmenting Images and Feature Descriptors..... 341

 Pipeline Stages and Operations 343

Mapping Operations to Resources	343
Criteria for Resource Assignments	344
Augmented Reality	345
Calibration and Ground Truth Data	346
Feature and Object Description	346
Overlays and Tracking	347
Pipeline Stages and Operations	348
Mapping Operations to Resources	348
Criteria for Resource Assignments	349
Acceleration Alternatives	350
Memory Optimizations	351
Minimizing Memory Transfers Between Compute Units	351
Memory Tiling	352
DMA, Data Copy, and Conversions	352
Register Files, Memory Caching, and Pinning	352
Data Structures, Packing, and Vector vs. Scatter-Gather Data Organization	353
Coarse-Grain Parallelism	353
Compute-Centric vs. Data-Centric	353
Threads and Multiple Cores	354
Fine-Grain Data Parallelism	354
SIMD, SIMT, and SPMD Fundamentals	355
Shader Kernel Languages and GPGPU	356
Advanced Instruction Sets and Accelerators	357
Vision Algorithm Optimizations and Tuning	358
Compiler And Manual Optimizations	359
Tuning	360
Feature Descriptor Retrofit, Detectors, Distance Functions	360

Boxlets and Convolution Acceleration 361

Data-Type Optimizations, Integer vs. Float..... 361

Optimization Resources 362

Summary..... 363

■ Appendix A: Synthetic Feature Analysis..... 365

Background Goals and Expectations..... 366

Test Methodology and Results 368

 Detector Parameters Are Not Tuned for the Synthetic Alphabets 369

 Expectations for Test Results 370

Summary of Synthetic Alphabet Ground Truth Images..... 370

 Synthetic Interest Point Alphabet 371

 Synthetic Corner Point Alphabet..... 371

 Synthetic Alphabet Overlays..... 371

Test 1: Synthetic Interest Point Alphabet Detection 372

 Annotated Synthetic Interest Point Detector Results 374

 Entire Images Available Online 375

Test 2: Synthetic Corner Point Alphabet Detection..... 383

 Annotated Synthetic Corner Point Detector Results..... 384

 Entire Images Available Online 384

Test 3: Synthetic Alphabets Overlaid on Real Images 393

 Annotated Detector Results on Overlay Images 393

Test 4: Rotational Invariance for Each Alphabet 394

 Methodology for Determining Rotational Invariance 394

Analysis of Results and Non-Repeatability Anomalies..... 398

 Caveats..... 398

 Non-Repeatability in Tests 1 and 2..... 399

 Other Non-Repeatability in Test 3..... 400

■ CONTENTS

Test Summary	400
Future Work	400
■ Appendix B: Survey of Ground Truth Datasets.....	401
■ Appendix C: Imaging and Computer Vision Resources.....	411
Commercial Products	411
Open Source.....	412
Organizations, Institutions, and Standards.....	415
Journals and Their Abbreviations.....	417
Conferences and Their Abbreviations.....	417
Online Resources	418
■ Appendix D: Extended SDM Metrics	419
■ Bibliography	437
Index.....	465

About the Author



Scott Krig is a pioneer in computer imaging, computer vision, and graphics visualization. He founded Krig Research in 1988 (krigresearch.com), providing the world's first imaging and vision systems based on high-performance engineering workstations, super-computers, and dedicated imaging hardware, serving customers worldwide in 25 countries. Scott has provided imaging and vision solutions around the globe, and has worked closely with many industries, including aerospace, military, intelligence, law enforcement, government research, and academic organizations.

More recently, Scott has worked for major corporations and startups serving commercial markets, solving problems in the areas of computer vision, imaging, graphics, visualization, robotics, process control, industrial automation, computer security, cryptography, and consumer applications of imaging and machine vision to PCs, laptops, mobile phones, and tablets. Most recently, Scott provided direction for Intel Corporation in the area of depth-sensing and computer vision methods for embedded systems and mobile platforms.

Scott is the author of many patent applications worldwide in the areas of embedded systems, imaging, computer vision, DRM, and computer security, and studied at Stanford.

Scott also enjoys acoustic guitar design and lutherie work, particularly 12-string acoustic guitars, as well as acoustic guitar composition and performance.

Acknowledgments

This book would not be as well thought out without the early technical feedback, conversations, and observations on very rough materials by Vadim Pizarevsky of ITSEEZ, who also is a major force behind the OpenCV foundation. Vadim brings vast and quantitative expertise in computer vision across a wide range of application domains. Thanks, Vadim.

Special thanks also go to Stuart Douglas at Intel Press for the commission to write this book, and for introductions to people at Apress. Also, special thanks to the key editors at Apress, including Melissa Maldonado, Mark Powers, Jeffrey Pepper, Steve Weiss, Robert Hutchinson, James Markham, and Carole Berglie for making this book a reality, and for adding value through the editorial process.

Introduction

Dirt. This is a jar of dirt.

Yes.

... Is the jar of dirt going to help?

If you don't want it, give it back.

—*Pirates Of The Carribean*, Jack Sparrow and Tia Dalma

This work focuses on a slice through the field - Computer Vision Metrics - from the view of feature description metrics, or how to describe, compute and design the macro-features and micro-features that make up larger objects in images. The focus is on the pixel-side of the vision pipeline, rather than the back-end training, classification, machine learning and matching stages. This book is suitable for reference, higher-level courses, and self-directed study in computer vision. The book is aimed at someone already familiar with computer vision and image processing; however, even those new to the field will find good introductions to the key concepts at a high level, via the ample illustrations and summary tables.

I view computer vision as a mathematical artform and its researchers and practitioners as artists. So, this book is more like a tour through an art gallery rather than a technical or scientific treatise. Observations are provided, interesting questions are raised, a vision taxonomy is suggested to draw a conceptual map of the field, and references are provided to dig deeper. This book is like an attempt to draw a map of the world centered around feature metrics, inaccurate and fuzzy as the map may be, with the hope that others will be inspired to expand the level of detail in their own way, better than what I, or even a few people, can accomplish alone. If I could have found a similar book covering this particular slice of subject matter, I would not have taken on the project to write this book.

What is not in the Book

Readers looking for computer vision “how-to” source code examples, tutorial discussions, performance analysis, and short-cuts will not find them here, and instead should consult the well-regarded <http://opencv.org> library resources, including many fine books, online resources, source code examples, and several blogs. There is nothing better than OpenCV for the hands-on practitioner. For this reason, this book steers a clear path around duplication of the “how-to” materials already provided by the OpenCV community and elsewhere, and instead provides a counterpoint discussion, including a comprehensive survey, analysis and taxonomy of methods. Also, do not expect all computer vision topics to be covered deeply with proofs and performance analysis,

since the bibliography references cover these matters quite well: for example, machine learning, training and classification methods are only lightly introduced, since the focus here is on the feature metrics.

In summary, this book is about the feature metrics, showing “what” methods practitioners are using, with detailed observations and analysis of “why” those methods work, with a bias towards raising questions via observations rather than providing too many answers. I like the questions best because good questions lead to many good answers, and each answer is often pregnant with more good questions...

This book is aimed at a survey level, with a taxonomy and analysis, so no detailed examples of individual use-cases or horse races between methods are included. However, much detail is provided in over 540+ bibliographic references to dig deeper into practical matters. Additionally, some “how-to” and “hands-on” resources are provided in Appendix C. And a little ‘perfunctory’ source code accompanying parts of this book is available online, for Appendix A covering the interest point detector evaluations for the synthetic interest point alphabets introduced in Chapter 7; and in Appendix D for extended SDM metrics covered in Chapter 3.

What is in the Book

Specifically, Chapter 1 provides preamble on 2d image formation and 3d depth imaging, and Chapter 2 promotes intelligent image pre-processing to enhance feature description. Chapters 3 through 6 form the core discussion on feature description, with an emphasis on local features. Global and regional metrics are covered in Chapter 3, feature descriptor concepts in Chapter 4, a vision taxonomy is suggested in Chapter 5, and local feature description is covered in Chapter 6. Ground truth data is covered in Chapter 7, and Chapter 8 discusses hypothetical vision pipelines and hypothetical optimizations from an engineering perspective, as a set of exercises to tie vision concepts together into real systems (coursework assignments can be designed to implement and improve the hypothetical examples in Chapter 8). A set of synthetic interest point alphabets is developed in Chapter 7, and ten common detectors are run against those alphabets, with the results provided in Appendix A. It is difficult to cleanly partition all topics in image processing and computer vision, so there is some overlap in the chapters. Also, there are many hybrids used in practice, so there’s inevitable overlap in the Chapter 5 vision taxonomy, and creativity always arrives on the horizon to find new and unexpected ways of using old methods. However, the taxonomy is a starting point and helped to guide the organization of the book.

Therefore, the main goal has been to survey and understand the range of methods used to describe features, without passing judgment on which methods are better. Some history is presented to describe why certain methods were developed, and what properties of invariance or performance were the goals, and we leave the claims to be proven by others, since “how” each method is implemented determines performance and accuracy, and “what” each method is tested against in terms of ground truth data really tells the rest of the story. If we can glean good ideas from the work of others, that is a measure of the success of their work.

Scope

For brevity's sake, I exclude a deep treatment of selected topics not directly related to the computer vision metrics themselves; this is an unusual approach, since computer vision discussions typically include a wider range of topics. Specifically, the topics not covered deeply here include statistical and machine learning, classification and training, feature database construction and optimization, and searching and sorting. Bibliography references are provided instead. Distance functions are discussed, since they are directly linked to the feature metric. (A future edition of this book may contain a deep dive into the statistical and machine learning side of computer vision, but not now.)

Terminology Caveat

Sometimes terminology in the literature does not agree when describing similar concepts. So in some cases, terminology is adopted in this work that is not standardized across independent research communities. In fact, some new and nonstandard terminology may be introduced here, possibly because the author is unaware of better existing terminology (perhaps some of the terminology introduced in this work will become standardized). Terminology divergence is most pronounced with regard to mathematical topics like clustering, regression, group distance, and error minimization, as well as for computer vision topics like keypoints, interest points, anchor points, and the like. The author recognizes that one is reluctant to change terminology, since so many concepts are learned based on the terminology. I recall a friend of mine, Homer Mead, chief engineer for the lunar rover and AWACS radar at Boeing, who sub-consciously refused to convert from using the older term condenser to use the newer term capacitor.

Inspiration comes from several sources, mostly the opportunity of pioneering: there is always some lack of clarity, structure and organization in any new field as the boundaries expand, so in this vast field the opportunity to explore is compelling: to map out structure and pathways of knowledge that others may follow to find new fields of study, create better markers along the way, and extend the pathways farther.

The inspiration for this book has come from conversations with a wide range of people over the years. Where did it all start? It began at Boeing in the early 1980s, while I was still in college. I was introduced to computer graphics in the Advanced Development Research labs where I worked, when the first computer-shaded 3D renderings of the space shuttle were made in raster form. At that time, mainly vector graphics machines were being used, like Evans & Sutherland Picture Systems, and eventually a BARCO frame buffer was added to the lab, and advanced raster computer renderings of shaded images from graphics models were pioneered by Jeff Lane and his group, as well as Loren Carpenter. Fractals, NURBS, and A-buffer techniques were a few of the methods developed in the labs, and the math of computer graphics, such as bi-cubic patches and bi-quintic patches, scared me away from graphics initially. But I was attracted to single pixels in the BARCO frame buffer, one pixel and line and frame at a time, since they seemed so intuitive and obvious. I initially pursued imaging and computer vision rather than all the computer graphics and associated math. However, it turned out that the computer vision and image processing math was far more diverse and equally complex anyway. Since then I have also spent considerable time in computer graphics. Back in the mid-1980s, Don Snow, my boss, who was co-founder and VP of research at Pacific

Western Systems and later at Applied Precision, asked me to analyze the View-PRB fixed-function hardware unit for pattern recognition to use for automatic wafer probing (in case we needed to build something like it ourselves) to locate patterns on wafers and align the machine for probing. Correlation was used for pattern matching, with a scale-space search method we termed “super-pixels.” The matching rate was four 32x32 patches per second over NTSC with sub-pixel accuracy, and I computed position, rotation, and offsets to align the wafer prober stage to prepare for wafer probing; we called this auto-align. I designed a pattern recognition servo system to locate the patterns with rotational accuracy of a few micro-radians, and positional accuracy of a fraction of a micron. In the later 1980s, I went to work for Mentor Graphics, and after several years I left the corporate R&D group reporting to the president Gerry Langelier to start a company, Krig Research, to focus on computer vision and imaging for high-end military and research customers based on expensive and now extinct workstations (SGI, Apollo, Sun... gone, all gone now...), and I have stayed interested ever since. Many things have changed in our industry; the software seems to all be free, and the hardware or SOC is almost free as well, so I am not sure how anyone can make any money at this anymore.

More recently, others have also provided inspiration. Thanks to Paul Rosin for synthetic images and organizational ideas. Thanks to Yann LeCun for providing key references into deep learning and convolutional networks, and thanks to Shree Nayar for permission to use a few images, and continuing to provide the computer vision community with inspiration via the Cave Research projects. And thanks to Luciano Oviedo for vast coverage of industry activity and strategy about where it is all going, and lively discussions.

Others, too many to list, have also added to my journey. And even though the conversations have sometimes been brief, or even virtual via email or SKYPE in many cases, the influence of their work and thinking has remained, so special thanks are due to several people who have provided comments to the manuscript or book outline, contributed images, or just plain inspiration they may not realize. Thank you to Rahul Suthankar, Alexandre Alahi for use of images and discussions; Steve Seitz, Bryan Russel, Liefeng Bo, and Xiaofeng Ren for deep-dive discussions about RGB-D computer vision and other research topics; Gutemberg Guerra-filho, Harsha Viswana, Dale Hitt, Joshua Gleason, Noah Snavely, Daniel Scharstein, Thomas Salmon, Richard Baraniuk, Carl Vodrick, Hervé Jégou, and Andrew Richardson; and also thanks for many interesting discussions on computer vision topics with several folks at Intel including Ofri Weschler, Hong Jiang, Andy Kuzma, Michael Jeronimo, Eli Turiel, and many others whom I have failed to mention.

Summary

In summary, my goal is to survey the methods people are using for feature description—the key metrics generated—and make it easier for anyone to understand the methods in practice, and how to evaluate the methods using the vision taxonomy and robustness criteria to get the results they are looking for, and find areas for extending the state of the art. And after hearing all the feedback from the first version of this work, I hope to create a second version that is even better.

Scott Krig
Anno Domini 2014