# Modeling and Optimization in Science and Technologies

Volume 12

The book series *Modeling and Optimization in Science and Technologies* (*MOST*) publishes basic principles as well as novel theories and methods in the fast-evolving field of modeling and optimization. Topics of interest include, but are not limited to: methods for analysis, design and control of complex systems, networks and machines; methods for analysis, visualization and management of large data sets; use of supercomputers for modeling complex systems; digital signal processing; molecular modeling; and tools and software solutions for different scientific and technological purposes. Special emphasis is given to publications discussing novel theories and practical solutions that, by overcoming the limitations of traditional methods, may successfully address modern scientific challenges, thus promoting scientific and technological progress. The series publishes monographs, contributed volumes and conference proceedings, as well as advanced textbooks. The main targets of the series are graduate students, researchers and professionals working at the forefront of their fields.

More information about this series at http://www.springer.com/series/10577

Shengrong Gong · Chunping Liu
Yi Ji · Baojiang Zhong · Yonggang Li
Husheng Dong

# Advanced Image and Video Processing Using MATLAB

Springer

Shengrong Gong
School of Computer Science
and Engineering
Changshu Institute of Technology
Changshu, Jiangsu, China

Chunping Liu
School of Computer Science
and Technology
Soochow University
Suzhou, Jiangsu, China

Yi Ji
School of Computer Science
and Technology
Soochow University
Suzhou, Jiangsu, China

Baojiang Zhong
School of Computer Science
and Technology
Soochow University
Suzhou, Jiangsu, China

Yonggang Li
College of Mathematics Physics
and Information Engineering
Jiaxing University
Jiaxing, Zhejiang, China

Husheng Dong
School of Computer Science
and Technology
Soochow University
Suzhou, Jiangsu, China

ISSN 2196-7326 ISSN 2196-7334 (electronic)
Modeling and Optimization in Science and Technologies
ISBN 978-3-319-77221-9 ISBN 978-3-319-77223-3 (eBook)
https://doi.org/10.1007/978-3-319-77223-3

Library of Congress Control Number: 2018948687



This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

# Preface

Digital image processing mainly focuses on the research of signal processing, such as image contrast adjustment, image coding, image denoising and filtering. It is different that Image analysis emphasizes describing images with symbolic representations, analysis, interpretation, and recognition. Along with the boom in artificial intelligent and deep learning, digital image processing is going deeper and more advanced. People start the researches in simulating the human vision to see, to understand, and even to explain the real world using three techniques: image segmentation, image analysis, and image understanding. The Image segmentation is to extract the features such as the edges and regions for image analyzing, recognition, and understanding. Image analysis is to extract intelligent information from underlying features and their relationship using mathematical models and Image processing techniques. Image analysis and image processing are closely related. Although there may be a certain degree of overlapping, they are different in essence. Therefore, image analysis is more related to pattern recognition and computer vision. It is generally used to analyze the underlying features and superstructures by some mathematical models. The researches of image analysis are mainly focused on content-based image retrieval, face recognition, emotion recognition, optical character recognition, handwriting recognition, biomedical image analysis, video object extraction. The Image understanding is to further understand the meanings and scenario explanations by researching the properties and relations of the features and objects. The objects for image understanding are symbols from description; the process is similar to human brain.

Corresponding to image analysis, video analysis is to analyze the video frames of surveillance camera using computer vision techniques. It is also able to filter the background such as wind, rain, snow, fallen leaves, birds, and floating flags. It is so called object tracking in complex background. Due to the variant illusion, motion, occlusion, color, and complex background, the difficulty of object detection and tracking algorithm design is increased.

The steps in image and video analysis mainly include segmentation, classification, and explanation. The classification process normally extracts the features by SIFT and LBP. With the use of deep learning techniques, people start using deep feature by extracting automatically for image classification, scenario classification, and behaviors analysis.

Our purpose in writing this book is to present advanced applications in image and video processing. We believed that this book is distinguished from other MATLAB-based fundamental textbooks which only introduces the basic functions such as the transform, enhancement, restoration, coding, and resizing of image. Our book emphasized the advanced applications such as image dehazing correction, image deraining correction, image stitching, image watermarking, visual object recognition, moving object tracking, dynamic scene classification, pedestrian re-identification, behavior analysis with deep learning, and so on.

The book is divided into three parts:

## Part I: The Basic Concepts

Chapter 1 briefly introduces the fundamental principles including the analysis techniques: scene segmentation, feature description, and object recognition. There are also some summaries about examples of advanced applications, such as image fusion, image inpainting, image stitching, image watermarking, object tracking, and pedestrian re-identification.

Chapter 2 introduces the functions of MATLAB toolboxes for image and video processing.

Chapter 3 presents the image and video segmentation methods of threshold, region-based, partial differential equation, clustering, graph theory, and cumulative difference-based motion region extraction.

Chapter 4 presents the feature extraction and representations, which includes Harris corner detection, SUSAN edge detection, the point feature detection algorithm SIFT and SURF.

## Part II: Advances in Image Processing

This part includes the image processing techniques such as image correction, image inpainting, image fusions, image stitching, image watermarking.

Chapter 5 firstly introduces three filters for image denoise and blurred functions. Then, it mainly introduces the correction techniques of image dehazing, image deraining, and text image feature correction.

Chapter 6 presents the image inpainting techniques including the principle, structure, algorithm, and some example codes.

Chapter 7 firstly introduces the fusions types and their schemes and then mentioned a very important method: wavelet transform for image fusion. Finally, it discusses the evaluation of image fusion objectively and subjectively.

Chapter 8 introduces the image stitching techniques such as region-based, feature-based, and feature point method. The SIFT and Harris corner detection algorithms are also introduced in this chapter.

Chapter 9 briefly introduces the image watermarking in three different transforms which are spatial-domain-based, DCT-based and DWT-based watermarking techniques.

Chapter 10 introduces the object recognition techniques including face recognition, facial expression, and image-to-character extraction and recognition.

## Part III: Advances in Video Processing and then Associated Chapters

Chapters 11–14 mainly introduce the video processing techniques of moving object tracking, dynamic scene classification based on TMBP, behavior recognition based on LDA topic model, person re-identification based on metric learning, lip recognition instance based on deep learning model, and deep CNN architecture for event recognition.

Chapter 11 introduces the object tracking techniques using Gaussian mixture model for background detection, and the RANSAC for feature points tracking. Further extend the mean-shift object tracking algorithm.

Chapter 12 introduces the dynamic scene classification and discusses the TMBP and LDA models for the classification.

Chapter 13 presents a person re-identification method by using the image understanding technique.

Chapter 14 presents the deep learning in image and video understanding.

For the convenience of the readers to evaluate the performance of the algorithms, we also give the common evaluation criteria in the appendix.

This book is written by Shengrong Gong, Chunping Liu, Yi Ji, Baojiang Zhong, Yonggang Li, Husheng Dong, Conghua Xie, Wei Pan, Yu xia, and Zhaohui Wang. Our M.Sc. researchers take participate in debugging most of the programs. They are Xinhua Dai, Ran Yan, Zongming Bao, and Pengcheng Zhou.

We gratefully acknowledge the professional suggestions of the reviewers and editors of Springer Press. We also thank to the support of Changshu Institute of Technology and Soochow University. We appreciate the support of National Natural Science Foundation of China (NSFC Grant No. 61170124, 61272258,

61301299), Integration of Cloud Computing and Big Data, Innovation of Science and Education (Grant No. 2017B03112), Provincial Natural Science Foundation of Jiangsu, China (Grant No. BK20151260, BK20151254), Six talent peaks project in Jiangsu Province, China (Grant No. DZXX-027).

Changshu, China — Shengrong Gong
Suzhou, China — Chunping Liu
Suzhou, China — Yi Ji
Suzhou, China — Baojiang Zhong
Jiaxing, China — Yonggang Li
Suzhou, China — Husheng Dong

# Contents

**Part I The Basic Concepts**

**1 Introduction** . . . . . 3
1.1 Basic Concepts and Terminology . . . . . 3
1.1.1 Digital Image and Digital Video . . . . . 3
1.1.2 Image Processing . . . . . 6
1.1.3 Image Analysis . . . . . 6
1.1.4 Video Analysis . . . . . 8
1.2 Image and Video Analysis . . . . . 9
1.2.1 Image and Video Scene Segmentation . . . . . 9
1.2.2 Image and Video Feature Description . . . . . 10
1.2.3 Object Recognition in Images/Videos . . . . . 12
1.2.4 Scene Description and Understanding . . . . . 13
1.3 Examples of Advanced Applications . . . . . 14
1.3.1 Image Correction . . . . . 14
1.3.2 Image Fusion . . . . . 15
1.3.3 Digital Image Inpainting . . . . . 15
1.3.4 Image Stitching . . . . . 16
1.3.5 Digital Watermarking . . . . . 17
1.3.6 Visual Object Recognition . . . . . 18
1.3.7 Object Tracking . . . . . 20
1.3.8 Dynamic Scene Classification . . . . . 21
1.3.9 Pedestrian Re-identification . . . . . 22
1.3.10 Lip Recognition in Video . . . . . 22
References . . . . . 23

**2 Matlab Functions of Image and Video** . . . . . 27
2.1 Introduction to MATLAB for Image and Video . . . . . 27
2.2 Basic Elements of MATLAB . . . . . 28

2.2.1 Working Environment . . . . . 28
2.2.2 Data Types . . . . . 29
2.2.3 Array and Matrix Indexing in MATLAB . . . . . 32
2.2.4 Standard Arrays . . . . . 34
2.2.5 Command-Line Operations . . . . . 34
2.3 Programming Tools: Scripts and Functions . . . . . 35
2.3.1 M-Files . . . . . 35
2.3.2 Operators . . . . . 36
2.3.3 Important Variables and Constants . . . . . 38
2.3.4 Number Representation . . . . . 38
2.3.5 Flow Control . . . . . 39
2.3.6 Input and Output . . . . . 41
2.4 Graphics and Visualization . . . . . 41
2.5 The Image Processing Toolbox . . . . . 46
2.5.1 The Image Processing Toolbox: An Overview . . . . . 46
2.5.2 Essential Functions and Features . . . . . 47
2.5.3 Displaying Information About an Image File . . . . . 52
2.5.4 Reading an Image File . . . . . 52
2.5.5 Data Classes and Data Conversions . . . . . 53
2.5.6 Displaying the Contents of an Image . . . . . 55
2.5.7 Exploring the Contents of an Image . . . . . 57
2.5.8 Writing the Resulting Image onto a File . . . . . 58
2.6 Video Processing in MATLAB . . . . . 58
2.6.1 Reading Video Files . . . . . 59
2.6.2 Processing Video Files . . . . . 59
2.6.3 Playing Video Files . . . . . 60
2.6.4 Writing Video Files . . . . . 61
2.6.5 Basic Digital Video Manipulation in MATLAB . . . . . 62
References . . . . . 63

**3 Image and Video Segmentation** . . . . . 65
3.1 Introduction . . . . . 65
3.2 Threshold Segmentation . . . . . 66
3.2.1 Global Threshold Image Segmentation . . . . . 68
3.2.2 Local Dynamic Threshold Segmentation . . . . . 69
3.3 Region-Based Segmentation . . . . . 74
3.3.1 Region Growing . . . . . 74
3.3.2 Region Splitting and Merging . . . . . 78
3.4 Segmentation Based on Partial Differential Equation . . . . . 88
3.5 Image Segmentation Based on Clustering . . . . . 94
3.6 Image Segmentation Method Based on Graph Theory . . . . . 97
3.6.1 Introduction . . . . . 97
3.6.2 GraphCut and Improved Image Segmentation Method . . . . . 99

3.7 Video Motion Region Extraction Method Based on Cumulative Difference . . . 107
References . . . 111

**4 Feature Extraction and Representation** . . . 113
4.1 Introduction . . . 113
4.2 Histogram-Based Features . . . 115
4.2.1 Grayscale Histogram . . . 115
4.2.2 Histograms of Oriented Gradients . . . 117
4.3 Texture Features . . . 121
4.3.1 Haralick Texture Descriptors . . . 122
4.3.2 Wavelet Texture Descriptors . . . 126
4.3.3 LBP Texture Descriptors . . . 131
4.4 Corner Feature Extraction . . . 135
4.4.1 Moravec Algorithm . . . 135
4.4.2 Harris Corner Detection Operator . . . 137
4.4.3 SUSAN Corner Detection Algorithm . . . 141
4.5 Local Invariant Feature Point Extraction . . . 144
4.5.1 Local Invariant Point Feature of SURF . . . 145
4.5.2 SIFT Scale-Invariant Feature Algorithm . . . 149
References . . . 158

**Part II Advances in Image Processing**

**5 Image Correction** . . . 161
5.1 Introduction . . . 161
5.2 Noise Reduction Using Spatial-Domain Techniques . . . 161
5.2.1 Selected Noise Probability Density Functions . . . 162
5.2.2 Filtering . . . 168
5.3 Image Deblurring . . . 173
5.3.1 The Restoration of Defocus Blurred Image . . . 174
5.3.2 Restoration of Motion Blurred Image . . . 176
5.4 Fisheye Distortion Correction Using Spherical Coordinates Model . . . 180
5.5 Skew Correction of Text Images . . . 186
5.5.1 Feature Analysis of Text Images . . . 187
5.5.2 The Basic Idea of Hough Transform . . . 187
5.5.3 The Implementation Steps of Text Images Skew Correction . . . 188
5.6 Image Dehazing Correction . . . 191
5.6.1 Single Image Dehazing . . . 191
5.6.2 Dark Channel Prior . . . 192
5.6.3 Implementation Steps of DCP . . . 194
5.6.4 Refine Transmission Map Using Soft Matting . . . 195

5.7 Image Deraining Correction . . . . . 200
5.7.1 Related Work . . . . . 200
5.7.2 Single Image De-rain with Deep Detail Network . . . . . 200
5.7.3 Implementation of Image Deraining with Deep Network . . . . . 203
References . . . . . 206

**6 Image Inpainting** . . . . . 209
6.1 Introduction . . . . . 209
6.1.1 Structure Oriented Image Inpainting Technology . . . . . 210
6.1.2 Texture-Based Image Inpainting Technology . . . . . 211
6.2 The Principle of Image Inpainting . . . . . 211
6.3 Variational PDE-Based Image Inpainting . . . . . 213
6.3.1 Image Inpainting Algorithm Based on Total Variational Model . . . . . 214
6.3.2 Image Inpainting Based on CDD Model . . . . . 219
6.4 Exemplar-Based Image Inpainting Algorithm . . . . . 222
References . . . . . 230

**7 Image Fusion** . . . . . 233
7.1 Introduction . . . . . 233
7.2 Fusion Categories . . . . . 234
7.2.1 Multi-view Fusion . . . . . 234
7.2.2 Multimodal Fusion . . . . . 236
7.2.3 Multi-temporal Fusion . . . . . 240
7.2.4 Multi-focus Fusion . . . . . 242
7.3 Image Fusion Schemes . . . . . 243
7.4 Image Fusion Using Wavelet Transform . . . . . 248
7.4.1 Basis of Wavelet Transform . . . . . 248
7.4.2 Discrete Dyadic Wavelet Transform of Image and Its Mallat Algorithm . . . . . 249
7.4.3 Steps of Implementation . . . . . 250
7.5 Region-Based Image Fusion . . . . . 253
7.5.1 Basic Framework of Regional Integration . . . . . 254
7.5.2 The Strategy of Regional Joint Representation . . . . . 255
7.5.3 The Rules of Fusion . . . . . 256
7.5.4 Wavelet Fusion of Regional Variance . . . . . 256
7.6 Image Fusion Using Fuzzy Dempster-Shafer Evidence Theory . . . . . 260
7.7 Image Quality and Fusion Evaluations . . . . . 263
7.7.1 Subjective Evaluation of Image Fusion . . . . . 264
7.7.2 Objective Evaluation of Image Fusion . . . . . 264
References . . . . . 268

**8 Image Stitching** . . . . . 271
8.1 Introduction . . . . . 271
8.2 Image Stitching Based on Region . . . . . 272
8.2.1 Image Stitching Based on Ratio Matching . . . . . 273
8.2.2 Image Stitching Based on Line and Plane Feature . . . . 276
8.2.3 Image Stitching Based on FFT . . . . . 283
8.3 Images Stitching Based on Feature Points . . . . . 290
8.3.1 SIFT Feature Points Detection . . . . . 290
8.3.2 Image Stitching Based on Harris Feature Points . . . . . 297
8.3.3 Auto-Sorting for Image Sequence . . . . . 304
8.3.4 Harris Point Registration Based on RANSAC Algorithm . . . . . 307
8.4 Panoramic Image Stitching . . . . . 320
References . . . . . 327

**9 Image Watermarking** . . . . . 329
9.1 Introduction . . . . . 329
9.2 Fragile Watermarking Based on Spatial Domain . . . . . 334
9.3 Robust Watermarking Based on DCT . . . . . 336
9.4 Semi-fragile Watermarking Based on DWT . . . . . 344
References . . . . . 349

**10 Visual Object Recognition** . . . . . 351
10.1 Face Recognition Based on Locality Preserving Projections . . . 351
10.2 Facial Expression Recognition Using PCA . . . . . 375
10.3 Extraction and Recognition of Characters in Pictures . . . . . 380
References . . . . . 387

**Part III Advances in Video Processing and then Associated Chapters**

**11 Visual Object Tracking** . . . . . 391
11.1 Adaptive Background Modeling by Using a Mixture of Gaussians . . . . . 391
11.2 Object Tracking Based on Ransac . . . . . 396
11.3 Object Tracking Based on MeanShift . . . . . 401
11.3.1 Description of the Object Model . . . . . 402
11.3.2 A Description of the Candidate Model . . . . . 402
11.3.3 Similarity Function . . . . . 403
11.3.4 Object Location . . . . . 403
11.4 Object Tracking Based on Particle Filter . . . . . 409
11.4.1 Prior Knowledge of the Goal . . . . . 410
11.4.2 System State Transition . . . . . 410
11.4.3 System Observation . . . . . 411
11.4.4 Posterior Probability Calculation . . . . . 412

11.4.5 Particle Resampling . . . . . 412
11.4.6 Implementation Steps . . . . . 413
11.5 Multiple Object Tracking . . . . . 418
References . . . . . 427

**12 Dynamic Scene Classification Based on Topic Models** . . . . . 429
12.1 Overview . . . . . 429
12.2 Introduction to the Topic Models . . . . . 430
12.2.1 LDA Model . . . . . 430
12.2.2 TMBP Model Based on Factor Graph . . . . . 433
12.2.3 TMBP Model Fusing Prior Knowledge . . . . . 436
12.3 Dynamic Scene Classification Based on TMBP . . . . . 439
12.4 Behavior Recognition Based on LDA Topic Model . . . . . 451

**13 Image Understanding-Person Re-identification** . . . . . 475
13.1 Introduction . . . . . 475
13.2 Person Re-ID Scenarios . . . . . 477
13.3 Methodology . . . . . 478
13.4 Public Datasets and Evaluation Metrics in Person Re-identification . . . . . 480
13.4.1 Public Datasets . . . . . 480
13.4.2 Evaluation Metrics . . . . . 483
13.5 Classic Feature Representations for Person Re-identification . . . . . 484
13.5.1 Salient Color Names . . . . . 484
13.5.2 Local Maximal Occurrence Representation . . . . . 487
13.6 An Example of Metric Learning Based Person Re-identification Method-XQDA . . . . . 501
References . . . . . 511

**14 Image and Video Understanding Based on Deep Learning** . . . . . 513
14.1 Introduction . . . . . 513
14.2 Model Analysis of CNN . . . . . 515
14.2.1 Basic Modules of CNN . . . . . 515
14.2.2 Convolution and Pooling . . . . . 515
14.2.3 Activation Function . . . . . 516
14.2.4 Softmax Classifier and Cost Function . . . . . 517
14.2.5 Learning Algorithm . . . . . 519
14.2.6 Dropout . . . . . 521
14.2.7 Batch Normalization . . . . . 522
14.3 Typical CNN Models . . . . . 522
14.3.1 LeNet . . . . . 522

14.3.2 AlexNet . . . . . 523
14.3.3 GoogLeNet . . . . . 524
14.3.4 VGGNet . . . . . 528
14.3.5 ResNet . . . . . 530
14.4 Deep Learning Model for Lip Recognition Instance . . . . . 531
14.4.1 Testing Dataset . . . . . 531
14.4.2 Deep Network Training . . . . . 532
14.4.3 Code Analysis . . . . . 536
14.5 Deep CNN Architecture for Event Recognition Instance . . . . . 539
14.5.1 Testing Dataset . . . . . 539
14.5.2 Deep Feature Extraction . . . . . 540
14.5.3 Spatial-Temporal Feature Fusion . . . . . 540
14.5.4 Fisher Vector Encoding . . . . . 541
14.5.5 Code Analysis . . . . . 542
References . . . . . 553

**Appendix: Common Evaluation Criterion** . . . . . 555

# About the Authors

**Shengrong Gong** received his M.S. from Harbin Institute of Technology in 1993 and his Ph.D. from Beihang University in 2001. He is the Dean of School of Computer Science and Engineering, Changshu Institute of Technology, and also a Professor and Doctoral Supervisor. His research interests are image and video processing, pattern recognition, and computer vision.

**Chunping Liu** received her Ph.D. in pattern recognition and artificial intelligence from Nanjing University of Science and Technology in 2002. She is now a Professor of School of Computer Science and Technology, Soochow University. Her research interests include computer vision, image analysis and recognition, in particular in the domains of visual saliency detection, object detection and recognition, and scene understanding.

**Yi Ji** received her M.S. from National University of Singapore, Singapore, and Ph.D. from INSA de Lyon, France. She is now an Associate Professor in School of Computer Science and Technology, Soochow University. Her research areas are 3D action recognition and complex scene understanding.

**Baojiang Zhong** received his B.S. in mathematics from Nanjing Normal University, China, in 1995, M.S. in mathematics, and Ph.D. in mechanical and electrical engineering from Nanjing University of Aeronautics and Astronautics (NUAA), China, in 1998 and 2006, respectively. From 1998 to 2009, he was on the Faculty of the Department of Mathematics of NUAA and reached the rank of Associate Professor. During 2007–2008, he was also a Research Scientist at the Temasek Laboratories, Nanyang Technological University, Singapore. In 2009, he joined the School of Computer Science and Technology, Soochow University, China, where he is currently a Full Professor. His research interests include computer vision, image processing, and numerical linear algebra.

**Yonggang Li** received his M.S. from Xi'an Polytechnic University in 2005. He is currently pursuing Ph.D. in School of Computer Science and Technology, Soochow University. He is a Lecturer of College of Mathematics, Physics and Information

Engineering, Jiaxing University. His research interests include computer vision, image and video processing, and pattern recognition.

**Husheng Dong** received his M.S. from School of Computer Science and Technology, Soochow University, in 2008, and he is pursuing Ph.D. currently. He is also a Lecturer of Suzhou Institute of Trade & Commerce. His research interest includes computer vision, image and video processing, and machine learning.