AN INTRODUCTION TO 3D COMPUTER VISION TECHNIQUES AND ALGORITHMS

Bogusław Cyganek
Department of Electronics, AGH University of Science and Technology, Poland

J. Paul Siebert
Department of Computing Science, University of Glasgow, Scotland, UK

WILEY
A John Wiley and Sons, Ltd., Publication

AN INTRODUCTION TO 3D COMPUTER VISION TECHNIQUES AND ALGORITHMS

AN INTRODUCTION TO 3D COMPUTER VISION TECHNIQUES AND ALGORITHMS

Bogusław Cyganek
Department of Electronics, AGH University of Science and Technology, Poland

J. Paul Siebert
Department of Computing Science, University of Glasgow, Scotland, UK

WILEY
A John Wiley and Sons, Ltd., Publication

This edition first published 2009
© 2009 John Wiley & Sons, Ltd

Registered office
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com.

The right of the author to be identified as the author of this work has been asserted in accordance with the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by the UK Copyright, Designs and Patents Act 1988, without the prior permission of the publisher.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The publisher is not associated with any product or vendor mentioned in this book. This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Library of Congress Cataloging-in-Publication Data

Cyganek, Boguslaw.
An introduction to 3D computer vision techniques and algorithms / by Boguslaw Cyganek and J. Paul Siebert.
p. cm.
Includes index.
ISBN 978-0-470-01704-3 (cloth)
1. Computer vision. 2. Three-dimensional imaging. 3. Computer algorithms. I. Siebert, J. Paul. II. Title
TA1634.C94 2008
006.3′7–dc22

2008032205

A catalogue record for this book is available from the British Library.

ISBN 978-0-470-01704-3

Set in 10/12pt Times by Aptara Inc., New Delhi, India.
Printed in Great Britain by CPI Antony Rowe, Chippenham, Wiltshire

To Magda, Nadia and Kamil
From Bogusław

To Sabina, Konrad and Gustav
From Paul

Contents

Preface **xv**

Acknowledgements **xvii**

Notation and Abbreviations **xix**

Part I 1

1 Introduction **3**
1.1 Stereo-pair Images and Depth Perception 4
1.2 3D Vision Systems 4
1.3 3D Vision Applications 5
1.4 Contents Overview: The 3D Vision Task in Stages 6

2 Brief History of Research on Vision **9**
2.1 Abstract 9
2.2 Retrospective of Vision Research 9
2.3 Closure 14
2.3.1 Further Reading 14

Part II 15

3 2D and 3D Vision Formation **17**
3.1 Abstract 17
3.2 Human Visual System 18
3.3 Geometry and Acquisition of a Single Image 23
3.3.1 Projective Transformation 24
3.3.2 Simple Camera System: the Pin-hole Model 24
3.3.2.1 Extrinsic Parameters 26
3.3.2.2 Intrinsic Parameters 27
3.3.3 Projective Transformation of the Pin-hole Camera 28
3.3.4 Special Camera Setups 29
3.3.5 Parameters of Real Camera Systems 30

3.4 Stereoscopic Acquisition Systems 31
3.4.1 Epipolar Geometry 31
3.4.1.1 Fundamental Matrix 34
3.4.1.2 Epipolar Lines and Epipoles 35
3.4.2 Canonical Stereoscopic System 36
3.4.3 Disparity in the General Case 38
3.4.4 Bifocal, Trifocal and Multifocal Tensors 39
3.4.5 Finding the Essential and Fundamental Matrices 41
3.4.5.1 Point Normalization for the Linear Method 44
3.4.5.2 Computing F in Practice 46
3.4.6 Dealing with Outliers 49
3.4.7 Catadioptric Stereo Systems 54
3.4.8 Image Rectification 55
3.4.9 Depth Resolution in Stereo Setups 59
3.4.10 Stereo Images and Reference Data 61
3.5 Stereo Matching Constraints 66
3.6 Calibration of Cameras 70
3.6.1 Standard Calibration Methods 71
3.6.2 Photometric Calibration 73
3.6.3 Self-calibration 73
3.6.4 Calibration of the Stereo Setup 74
3.7 Practical Examples 75
3.7.1 Image Representation and Basic Structures 75
3.7.1.1 Computer Representation of Pixels 76
3.7.1.2 Representation of Images 78
3.7.1.3 Image Operations 87
3.8 Appendix: Derivation of the Pin-hole Camera Transformation 91
3.9 Closure 93
3.9.1 Further Reading 93
3.9.2 Problems and Exercises 94

4 Low-level Image Processing for Image Matching 95
4.1 Abstract 95
4.2 Basic Concepts 95
4.2.1 Convolution and Filtering 95
4.2.2 Filter Separability 97
4.3 Discrete Averaging 99
4.3.1 Gaussian Filter 100
4.3.2 Binomial Filter 101
4.3.2.1 Specification of the Binomial Filter 101
4.3.2.2 Spectral Properties of the Binomial Filter 102
4.4 Discrete Differentiation 105
4.4.1 Optimized Differentiating Filters 105
4.4.2 Savitzky–Golay Filters 108
4.4.2.1 Generation of Savitzky–Golay Filter Coefficients 114

4.5 Edge Detection 115
4.5.1 Edges from Signal Gradient 117
4.5.2 Edges from the Savitzky–Golay Filter 119
4.5.3 Laplacian of Gaussian 120
4.5.4 Difference of Gaussians 126
4.5.5 Morphological Edge Detector 127
4.6 Structural Tensor 127
4.6.1 Locally Oriented Neighbourhoods in Images 128
4.6.1.1 Local Neighbourhood with Orientation 130
4.6.1.2 Definition of a Local Neighbourhood of Pixels 130
4.6.2 Tensor Representation of Local Neighbourhoods 133
4.6.2.1 2D Structural Tensor 136
4.6.2.2 Computation of the Structural Tensor 140
4.6.3 Multichannel Image Processing with Structural Tensor 143
4.7 Corner Detection 144
4.7.1 The Most Common Corner Detectors 144
4.7.2 Corner Detection with the Structural Tensor 149
4.8 Practical Examples 151
4.8.1 C++ Implementations 151
4.8.1.1 Convolution 151
4.8.1.2 Implementing the Structural Tensor 155
4.8.2 Implementation of the Morphological Operators 157
4.8.3 Examples in Matlab: Computation of the SVD 161
4.9 Closure 162
4.9.1 Further Reading 163
4.9.2 Problems and Exercises 163

5 Scale-space Vision 165
5.1 Abstract 165
5.2 Basic Concepts 165
5.2.1 Context 165
5.2.2 Image Scale 166
5.2.3 Image Matching Over Scale 166
5.3 Constructing a Scale-space 168
5.3.1 Gaussian Scale-space 168
5.3.2 Differential Scale-space 170
5.4 Multi-resolution Pyramids 172
5.4.1 Introducing Multi-resolution Pyramids 172
5.4.2 How to Build Pyramids 175
5.4.3 Constructing Regular Gaussian Pyramids 175
5.4.4 Laplacian of Gaussian Pyramids 177
5.4.5 Expanding Pyramid Levels 178
5.4.6 Semi-pyramids 179
5.5 Practical Examples 181
5.5.1 C++ Examples 181
5.5.1.1 Building the Laplacian and Gaussian Pyramids in C++ 181

5.5.2 Matlab Examples 186
5.5.2.1 Building the Gaussian Pyramid in Matlab 190
5.5.2.2 Building the Laplacian of Gaussians Pyramid in Matlab 190
5.6 Closure 191
5.6.1 Chapter Summary 191
5.6.2 Further Reading 191
5.6.3 Problems and Exercises 192

6 Image Matching Algorithms 193
6.1 Abstract 193
6.2 Basic Concepts 193
6.3 Match Measures 194
6.3.1 Distances of Image Regions 194
6.3.2 Matching Distances for Bit Strings 198
6.3.3 Matching Distances for Multichannel Images 199
6.3.3.1 Statistical Distances 201
6.3.4 Measures Based on Theory of Information 202
6.3.5 Histogram Matching 205
6.3.6 Efficient Computations of Distances 206
6.3.7 Nonparametric Image Transformations 209
6.3.7.1 Reduced Census *Coding* 212
6.3.7.2 Sparse Census *Relations* 214
6.3.7.3 Fuzzy Relationships Among Pixels 215
6.3.7.4 Implementation of Nonparametric Image Transformations 216
6.3.8 Log-polar Transformation for Image Matching 218
6.4 Computational Aspects of Matching 222
6.4.1 Occlusions 222
6.4.2 Disparity Estimation with Subpixel Accuracy 224
6.4.3 Evaluation Methods for Stereo Algorithms 226
6.5 Diversity of Stereo Matching Methods 229
6.5.1 Structure of Stereo Matching Algorithms 233
6.5.1.1 Aggregation of the Cost Values 234
6.5.1.2 Computation of the Disparity Map 235
6.5.1.3 Disparity Map Postprocessing 237
6.6 Area-based Matching 238
6.6.1 Basic Search Approach 239
6.6.2 Interpreting Match Cost 241
6.6.3 Point-oriented Implementation 245
6.6.4 Disparity-oriented Implementation 250
6.6.5 Complexity of Area-based Matching 256
6.6.6 Disparity Map Cross-checking 257
6.6.7 Area-based Matching in Practice 259
6.6.7.1 Intensity Matching 260
6.6.7.2 Area-based Matching in Nonparametric Image Space 260
6.6.7.3 Area-based Matching with the Structural Tensor 262

6.7 Area-based Elastic Matching 273
6.7.1 Elastic Matching at a Single Scale 273
6.7.1.1 Disparity Match Range 274
6.7.1.2 Search and Subpixel Disparity Estimation 275
6.7.2 Elastic Matching Concept 278
6.7.3 Scale-based Search 280
6.7.4 Coarse-to-fine Matching Over Scale 283
6.7.5 Scale Subdivision 284
6.7.6 Confidence Over Scale 285
6.7.7 Final Multi-resolution Matcher 286
6.8 Feature-based Image Matching 288
6.8.1 Zero-crossing Matching 289
6.8.2 Corner-based Matching 292
6.8.3 Edge-based Matching: The Shirai Method 295
6.9 Gradient-based Matching 296
6.10 Method of Dynamic Programming 298
6.10.1 Dynamic Programming Formulation of the Stereo Problem 301
6.11 Graph Cut Approach 306
6.11.1 Graph Cut Algorithm 306
6.11.1.1 Graphs in Computer Vision 309
6.11.1.2 Optimization on Graphs 310
6.11.2 Stereo as a Voxel Labelling Problem 311
6.11.3 Stereo as a Pixel Labelling Problem 312
6.12 Optical Flow 314
6.13 Practical Examples 318
6.13.1 Stereo Matching Hierarchy in C++ 318
6.13.2 Log-polar Transformation 319
6.14 Closure 321
6.14.1 Further Reading 321
6.14.2 Problems and Exercises 322

7 Space Reconstruction and Multiview Integration 323
7.1 Abstract 323
7.2 General 3D Reconstruction 323
7.2.1 Triangulation 324
7.2.2 Reconstruction up to a Scale 325
7.2.3 Reconstruction up to a Projective Transformation 327
7.3 Multiview Integration 329
7.3.1 Implicit Surfaces and Marching Cubes 330
7.3.1.1 Range Map Pre-segmentation 331
7.3.1.2 Volumetric Integration Algorithm Overview 332
7.3.1.3 Hole Filling 332
7.3.1.4 Marching Cubes 333
7.3.1.5 Implementation Considerations 338
7.3.2 Direct Mesh Integration 338

7.4 Closure 342
7.4.1 Further Reading 342

8 Case Examples **343**
8.1 Abstract 343
8.2 3D System for Vision-Impaired Persons 343
8.3 Face and Body Modelling 345
8.3.1 Development of Face and Body Capture Systems 345
8.3.2 Imaging Resolution, 3D Resolution and Implications for Applications 346
8.3.3 3D Capture and Analysis Pipeline for Constructing Virtual Humans 350
8.4 Clinical and Veterinary Applications 352
8.4.1 Development of 3D Clinical Photography 352
8.4.2 Clinical Requirements for 3D Imaging 353
8.4.3 Clinical Assessment Based on 3D Surface Anatomy 353
8.4.4 Extraction of Basic 3D Anatomic Measurements 354
8.4.5 Vector Field Surface Analysis by Means of Dense Correspondences 357
8.4.6 Eigenspace Methods 359
8.4.7 Clinical and Veterinary Examples 362
8.4.8 Multimodal 3D Imaging 367
8.5 Movie Restoration 370
8.6 Closure 374
8.6.1 Further Reading 374

Part III 375

9 Basics of the Projective Geometry **377**
9.1 Abstract 377
9.2 Homogeneous Coordinates 377
9.3 Point, Line and the Rule of Duality 379
9.4 Point and Line at Infinity 380
9.5 Basics on Conics 382
9.5.1 Conics in $\wp^2$ 382
9.5.1.1 The Dual Conic 383
9.5.1.2 Circular Points 383
9.5.2 Conics in $\wp^2$ 384
9.5.2.1 The Absolute Conic 384
9.5.2.2 The Dual Absolute Conic 385
9.6 Group of Projective Transformations 385
9.6.1 Projective Base 385
9.6.2 Hyperplanes 386
9.6.3 Projective Homographies 386
9.7 Projective Invariants 387
9.8 Closure 388
9.8.1 Further Reading 389

10 Basics of Tensor Calculus for Image Processing 391
10.1 Abstract 391
10.2 Basic Concepts 391
10.2.1 Linear Operators 392
10.2.2 Change of Coordinate Systems: Jacobians 393
10.3 Change of a Base 394
10.4 Laws of Tensor Transformations 396
10.5 The Metric Tensor 397
10.5.1 Covariant and Contravariant Components in a Curvilinear Coordinate System 397
10.5.2 The First Fundamental Form 399
10.6 Simple Tensor Algebra 399
10.6.1 Tensor Summation 399
10.6.2 Tensor Product 400
10.6.3 Contraction and Tensor Inner Product 400
10.6.4 Reduction to Principal Axes 400
10.6.5 Tensor Invariants 401
10.7 Closure 401
10.7.1 Further Reading 401

11 Distortions and Noise in Images 403
11.1 Abstract 403
11.2 Types and Models of Noise 403
11.3 Generating Noisy Test Images 405
11.4 Generating Random Numbers with Normal Distributions 407
11.5 Closure 408
11.5.1 Further Reading 408

12 Image Warping Procedures 409
12.1 Abstract 409
12.2 Architecture of the Warping System 409
12.3 Coordinate Transformation Module 410
12.3.1 Projective and Affine Transformations of a Plane 410
12.3.2 Polynomial Transformations 411
12.3.3 Generic Coordinates Mapping 412
12.4 Interpolation of Pixel Values 412
12.4.1 Bilinear Interpolation 412
12.4.2 Interpolation of Nonscalar-Valued Pixels 414
12.5 The Warp Engine 414
12.6 Software Model of the Warping Schemes 415
12.6.1 Coordinate Transformation Hierarchy 415
12.6.2 Interpolation Hierarchy 416
12.6.3 Image Warp Hierarchy 416
12.7 Warp Examples 419
12.8 Finding the Linear Transformation from Point Correspondences 420
12.8.1 Linear Algebra on Images 424

12.9 Closure 427
12.9.1 Further Reading 428

13 Programming Techniques for Image Processing and Computer Vision 429
13.1 Abstract 429
13.2 Useful Techniques and Methodology 430
13.2.1 Design and Implementation 430
13.2.1.1 Comments and Descriptions of 'Ideas' 430
13.2.1.2 Naming Conventions 431
13.2.1.3 Unified Modelling Language (UML) 431
13.2.2 Template Classes 436
13.2.2.1 Expression Templates 437
13.2.3 Asserting Code Correctness 438
13.2.3.1 Programming by Contract 438
13.2.4 Debugging Issues 440
13.3 Design Patterns 441
13.3.1 Template Function Objects 441
13.3.2 Handle-body or Bridge 442
13.3.3 Composite 445
13.3.4 Strategy 447
13.3.5 Class Policies and Traits 448
13.3.6 Singleton 450
13.3.7 Proxy 450
13.3.8 Factory Method 451
13.3.9 Prototype 452
13.4 Object Lifetime and Memory Management 453
13.5 Image Processing Platforms 455
13.5.1 Image Processing Libraries 455
13.5.2 Writing Software for Different Platforms 455
13.6 Closure 456
13.6.1 Further Reading 456

14 Image Processing Library 457

References 459

Index 475

Preface

Recent decades have seen rapidly growing research in many areas of computer science, including computer vision. This comes from the natural interest of researchers as well as demands from industry and society for qualitatively new features to be afforded by computers. One especially desirable capability would be automatic reconstruction and analysis of the surrounding 3D environment and recognition of objects in that space. Effective 3D computer vision methods and implementations would open new possibilities such as automatic navigation of robots and vehicles, scene surveillance and monitoring (which allows automatic recognition of unexpected behaviour of people or other objects, such as cars in everyday traffic), medical reasoning, remote surgery and many, many more.

This book is a result of our long fascination with computers and vision algorithms. It started many years ago as a set of short notes with the only purpose 'to remember this or that' or to have a kind of 'short reference' just for ourselves. However, as this diary grew with the years we decided to make it available to other people. We hope that it was a good decision! It is our hope that this book facilitates access to this enthralling area, especially for students and young researchers. Our intention is to provide a very concise, though as far as possible complete, overview of the basic concepts of 2D and 3D computer vision. However, the best way to get into the field is to try it oneself! Therefore, in parallel with explaining basic concepts, we provide also a basic programming framework with the hope of making this process easier. We greatly encourage the reader to take the next step and try the techniques in practice.

Bogusław Cyganek, Kraków, Poland
J. Paul Siebert, Glasgow, UK

Acknowledgements

We would like to express our gratitude to all the people who helped in the preparation of this book!

In particular, we are indebted to the whole Wiley team who helped in the preparation of the manuscript. In this group special thanks go to Simone Taylor who believed in this project and made it happen. We would also like to express our gratitude to Sian Andrews, Laura Bell, Liz Benson, Emily Bone, Lucy Bryan, Kate Griffiths, Wendy Hunter, Alex King, Erica Peters, Kathryn Sharples, and Nicky Skinner.

We are also very grateful to the individuals and organizations who agreed to the use of their figures in the book. These are Professor Yuichi Ohta from Tsukuba University, as well as Professor Ryszard Szeliski from Microsoft Research. Likewise we would like to thank Dimensional Imaging Ltd. and Precision 3D Ltd. for use of their images. In this respect we would also like to express our gratitude to Springer Science and Business Media, IEEE Computer Society Press, the IET, Emerald Publishing, the ACM, Maney Publishing and Elsevier Science.

We would also like to thank numerous colleagues from the AGH University of Science and Technology in Kraków. We owe a special debt of gratitude to Professor Ryszard Tadeusiewicz and Professor Kazimierz Wiatr, as well as to Lidia Krawentek for their encouragement and continuous support.

We would also like to thank members of the former Turing Institute in Glasgow (Dr Tim Niblett, Joseph Jin, Dr Peter Mowforth, Dr Colin Urquhart and also Arthur van Hoff) as well as members of the Computer Vision and Graphics Group in the Department of Computing Science, University of Glasgow, for access to and use of their research material (Dr John Patterson, Dr Paul Cockshott, Dr Xiangyang Ju, Dr Yijun Xiao, Dr Zhili Mao, Dr Zhifang Mao (posthumously), Dr J.C Nebel, Dr Tim Boyling, Janet Bowman, Susanne Oehler, Stephen Marshall, Don Whiteford and Colin McLaren). Similarly we would like to thank our collaborators in the Glasgow Dental Hospital and School (Professor Khursheed Moos, Professor Ashraf Ayoub and Dr Balvinder Khambay), Canniesburn Plastic Surgery Unit (Mr Arup Ray), Glasgow, the Department of Statistics (Professor Adrian Bowman and Dr Mitchum Bock), Glasgow University, Professor Donald Hadley, Institute of Neurological Sciences, Southern General Hospital, Glasgow, and also those colleagues formerly at the Silsoe Research Institute (Dr Robin Tillett, Dr Nigel McFarlane and Dr Jerry Wu), Silsoe, UK.

Special thanks are due to Dr Sumitha Balasuriya for use of his Matlab codes and graphs. Particular thanks are due to Professor "Keith" van Rijsbergen and Professor Ray Welland without whose support much of the applied research we report would not have been possible.

We wish to express our special thanks and gratitude to Steve Brett from Pandora Inc. for granting rights to access their software platform.

Some parts of the research for which results are provided in this book were possible due to financial support of the European Commission under RACINE-S (IST-2001-37117) and IP-RACINE (IST-2-511316-IP) as well as Polish funds for scientific research in 2007–2008. Research described in these pages has also been funded by the UK DTI and the EPSRC & BBSRC funding councils, the Chief Scientist Office (Scotland), Wellcome Trust, Smith's Charity, the Cleft Lip and Palate Association, the National Lottery (UK) and the Scottish Office. Their support is greatly appreciated.

Finally, we would like to thank Magda and Sabina for their encouragement, patience and understanding over the three-year period it took to write this book.

Notation and Abbreviations

$I_k(x, y)$	Intensity value of a k-th image at a point with local image coordinates (x, y)
$\overline{I_k(x, y)}$	Average intensity value of a k-th image at a point with local image coordinates (x, y)
$\mathbf{I}$	Identity matrix; image treated as a matrix
$\mathbf{P}$	A vector (a point), matrix, tensor, etc.
$T[\mathbf{I}, \mathbf{P}]$	The Census transformation T for a pixel $\mathbf{P}$ in the image $\mathbf{I}$
i, j	Free coordinates
d_x, d_y	Displacements (offset) in the x and y directions
$D(\mathbf{p}_l, \mathbf{p}_r)$	Disparity between points $\mathbf{p}_l$ and $\mathbf{p}_r$
$\mathbf{D}$	Disparity map (a matrix)
$U(x, y)$	Local neighbourhood of pixels around a point (x, y)
$\mathbf{O}_c$	Optical centre point
$\mathbf{P}_c = [X_c, Y_c, Z_c]^T$	Coordinates of a 3D point in the camera coordinate system
$\mathbf{\Pi}$	Camera plane; a projective plane
$\mathbf{o} = (o_x, o_y)$	Central point of a camera plane
f	Focus length of a camera
b	Base line in a stereo system (a distance between cameras)
h_x, h_y	Physical horizontal and vertical dimensions of a pixel
$\mathbf{P} = [X, Y, Z]^T$	3D point and its coordinates
$\wp^n$	N-dimensional projective space
$\mathbf{P} = [X_h, Y_h, Z_h, 1]^T$	Homogenous coordinates of a point
$\mathbf{M}$	Camera matrix
$\mathbf{M}_\mathrm{i}$	Intrinsic parameters of a camera
$\mathbf{M}_\mathrm{e}$	Extrinsic parameters of a camera
$\mathbf{E}$	Essential matrix.
$\mathbf{F}$	Fundamental matrix.
$\mathbf{e}_i$	Epipole in an i-th image
SAD	Sum of absolute differences
SSD	Sum of squared differences
ZSAD	Zero-mean sum of absolute differences
ZSSD	Zero-mean sum of squared differences
ZSSD-N	Zero-mean sum of squared differences, normalized

SCP	Sum of cross products
SCP-N	Sum of cross products, normalized
RMS	Root mean square
RMSE	Root mean square error
<Lxx, Lyy>	Code lines from a line *Lxx* to *Lyy*
HVS	Human Visual System
SDK	Software Development Kit
$\wedge$	logical 'and'
$\vee$	logical 'or'
LRC	Left-right checking (cross-checking)
OCC	Occlusion constraint
ORD	Point ordering constraint
BMD	Bimodality rule
MGJ	Match goodness jumps
NM	Null method
GT RMS	Ground-truth RMS
WTA	Winner-takes-all
*	Convolution operator

Part I

1

Introduction

The purpose of this text on stereo-based imaging is twofold: it is to give students of computer vision a thorough grounding in the image analysis and projective geometry techniques relevant to the task of recovering three-dimensional (3D) surfaces from stereo-pair images; and to provide a complete reference text for professional researchers in the field of computer vision that encompasses the fundamental mathematics and algorithms that have been applied and developed to allow 3D vision systems to be constructed.

Prior to reviewing the contents of this text, we shall set the context of this book in terms of the underlying objectives and the explanation and design of 3D vision systems. We shall also consider briefly the historical context of optics and vision research that has led to our contemporary understanding of 3D vision.

Here we are specifically considering 3D vision systems that base their operation on acquiring stereo-pair images of a scene and then decoding the depth information *implicitly* captured within the stereo-pair as parallaxes, i.e. relative displacements of the contents of one of the images of the stereo-pair with respect to the other image. This process is termed *stereo-photogrammetry*, i.e. measurement from stereo-pair images. For readers with normal functional binocular vision, the everyday experience of observing the world with both of our eyes results in the perception of the relative distance (depth) to points on the surfaces of objects that enter our field of view. For over a hundred years it has been possible to configure a stereo-pair of cameras to capture stereo-pair images, in a manner analogous to mammalian binocular vision, and thereafter view the developed photographs to observe a miniature 3D scene by means of a stereoscope device (used to present the left and right images of the captured stereo-pair of photographs to the appropriate eye). However, in this scenario it is the brain of the observer that must decode the depth information locked within the stereo-pair and thereby experience the perception of depth. In contrast, in this book we shall present underlying mechanisms by which a computer program can be devised to analyse digitally formatted images captured by a stereo-pair of cameras and thereby recover an *explicit* measurement of distances to points *sampling* surfaces in the imaged field of view. Only by explicitly recovering depth estimates does it become possible to undertake useful tasks such as 3D measurement or reverse engineering of object surfaces as elaborated below. While the science of stereo-photogrammetry is a well-established field and it has indeed been possible to undertake 3D

An Introduction to 3D Computer Vision Techniques and Algorithms Bogusław Cyganek and J. Paul Siebert
© 2009 John Wiley & Sons, Ltd

measurement by means of stereo-pair images using a manually operated measurement device (the stereo-comparator) since the beginning of the twentieth century, we present fully automatic approaches for 3D imaging and measurement in this text.

1.1 Stereo-pair Images and Depth Perception

To appreciate the structure of 3D vision systems based on processing stereo-pair images, it is first necessary to grasp, at least in outline, the most basic principles involved in the formation of stereo-pair images and their subsequent analysis. As outlined above, when we observe a scene with both eyes, an image of the scene is formed on the retina of each eye. However, since our eyes are horizontally displaced with respect to each other, the images thus formed are not identical. In fact this stereo-pair of retinal images contains slight displacements between the relative locations of local parts of the image of the scene with respect to each image of the pair, depending upon how close these local scene components are to the point of *fixation* of the observer's eyes. Accordingly, it is possible to reverse this process and deduce how far away scene components were from the observer according to the magnitude and direction of the parallaxes within the stereo-pairs when they were captured. In order to accomplish this task two things must be determined: firstly, those local parts of one image of the stereo-pair that match the corresponding parts in the other image of the stereo-pair, in order to find the local parallaxes; secondly, the precise geometric properties and configuration of the eyes, or cameras. Accordingly, a process of *calibration* is required to discover the requisite geometric information to allow the imaging process to be inverted and relative distances to surfaces observed in the stereo-pair to be recovered.

1.2 3D Vision Systems

By definition, a stereo-photogrammetry-based 3D vision system will require stereo-pair image acquisition hardware, usually connected to a computer hosting software that automates acquisition control. Multiple stereo-pairs of cameras might be employed to allow all-round coverage of an object or person, e.g. in the context of whole-body scanners. Alternatively, the object to be imaged could be mounted on a computer-controlled turntable and overlapping stereo-pairs captured from a fixed viewpoint for different turntable positions. Accordingly, sequencing capture and image download from multiple cameras can be a complex process, and hence the need for a computer to automate this process.

The stereo-pair acquisition process falls into two categories, active illumination and passive illumination. Active illumination implies that some form of pattern is projected on to the scene to facilitate finding and disambiguating parallaxes (also termed *correspondences* or *disparities*) between the stereo-pair images. Projected patterns often comprise grids or stripes and sometimes these are even colour coded. In an alternative approach, a random speckle texture pattern is projected on to the scene in order to augment the texture already present on imaged surfaces. Speckle projection can also guarantee that that imaged surfaces appear to be randomly textured and are therefore locally uniquely distinguishable and hence able to be matched successfully using certain classes of image matching algorithm. With the advent of 'high-resolution' digital cameras the need for pattern projection has been reduced, since the surface texture naturally present on materials, having even a matte finish, can serve to facilitate

matching stereo-pairs. For example, stereo-pair images of the human face and body can be matched successfully using ordinary studio flash illumination when the pixel sampling density is sufficient to resolve the natural texture of the skin, e.g. skin-pores. A camera resolution of approximately 8–13M pixels is adequate for stereo-pair capture of an area corresponding to the adult face or half-torso.

The acquisition computer may also host the principal 3D vision software components:

- An image matching algorithm to find correspondences between the stereo-pairs.
- Photogrammetry software that will perform system calibration to recover the geometric configuration of the acquisition cameras and perform 3D point reconstruction in world coordinates.
- 3D surface reconstruction software that builds complete manifolds from 3D *point-clouds* captured by each imaging stereo-pair.

3D visualisation facilities are usually also provided to allow the reconstructed surfaces to be displayed, often *draped* with an image to provide a *photorealistic* surface model. At this stage the 3D shape and surface appearance of the imaged object or scene has been captured in explicit digital metric form, ready to feed some subsequent application as described below.

1.3 3D Vision Applications

This book has been motivated in part by the need to provide a manual of techniques to serve the needs of the computer vision practitioner who wishes to construct 3D imaging systems configured to meet the needs of practical applications. A wide variety of applications are now emerging which rely on the fast, efficient and low-cost capture of 3D surface information. The traditional role for image-based 3D surface measurement has been the reserve of *close-range* photogrammetry systems, capable of recovering surface measurements from objects in the range of a few tens of millimetres to a few metres in size. A typical example of a classical close-range photogrammetry task might comprise surface measurement for manufacturing quality control, applied to high-precision engineered products such as aircraft wings.

Close-range video-based photogrammetry, having a lower spatial resolution than traditional plate-camera film-based systems, initially found a niche in imaging the human face and body for clinical and creative media applications. 3D clinical photographs have the potential to provide quantitative measurements that reduce subjectivity in assessing the surface anatomy of a patient (or animal) before and after surgical intervention by providing numeric, possibly automated, scores for the shape, symmetry and longitudinal change of anatomic structures. Creative media applications include whole-body 3D imaging to support creation of human avatars of specific individuals, for 3D gaming and cine special effects requiring virtual actors. Clothing applications include body or foot scanning for the production of custom clothing and shoes or as a means of sizing customers accurately. An innovative commercial application comprises a ‘virtual catwalk’ to allow customers to visualize themselves in clothing prior to purchasing such goods on-line via the Internet.

There are very many more emerging uses for 3D imaging beyond the above and commercial ‘reverse engineering’ of premanufactured goods. 3D vision systems have the potential to revolutionize autonomous vehicles and the capabilities of robot vision systems. Stereo-pair cameras could be mounted on a vehicle to facilitate autonomous navigation or configured

within a robot workcell to endow a 'blind' pick-and-place robot, both object recognition capabilities based on 3D cues and simultaneously 3D spatial quantification of object locations in the workspace.

1.4 Contents Overview: The 3D Vision Task in Stages

The organization of this book reflects the underlying principles, structural components and uses of 3D vision systems as outlined above, starting with a brief historical view of vision research in Chapter 2. We deal with the basic existence proof that binocular 3D vision is possible, in an overview of the human visual system in Chapter 3. The basic projective geometry techniques that underpin 3D vision systems are also covered here, including the geometry of monocular and binocular image formation which relates how binocular parallaxes are produced in stereo-pair images as a result of imaging scenes containing variation in depth. Camera calibration techniques are also presented in Chapter 3, completing the introduction of the role of image formation and geometry in the context of 3D vision systems.

We deal with fundamental 2D image analysis techniques required to undertake image filtering and feature detection and localization in Chapter 4. These topics serve as a precursor to perform image matching, the process of detecting and quantifying parallaxes between stereo-pair images, a prerequisite to recovering depth information. In Chapter 5 the issue of spatial scale in images is explored, namely how to structure algorithms capable of efficiently processing images containing structures of varying scales which are unknown in advance. Here the concept of an image *scale-space* and the *multi-resolution image pyramid* data structure is presented, analysed and explored as a precursor to developing matching algorithms capable of operating over a wide range of visual scales. The core algorithmic issues associated with stereo-pair image matching are contained in Chapter 6 dealing with distance measures for comparing image patches, the associated parametric issues for matching and an in-depth analysis of area-based matching over scale-space within a practical matching algorithm. Feature-based approaches to matching are also considered and their combination with area-based approaches. Then two solutions to the stereo problem are discussed: the first, based on the *dynamic programming*, and the second one based on the *graph cuts* method. The chapter ends with discussion of the *optical flow* methods which allow estimation of local displacements in a sequence of images.

Having dealt with the recovery of disparities between stereo-pairs, we progress logically to the recovery of 3D surface information in Chapter 7. We consider the process of *triangulation* whereby 3D points in world coordinates are computed from the disparities recovered in the previous chapter. These 3D points can then be organized into surfaces represented by *polygonal meshes* and the *3D point-clouds* recovered from *multi-view* systems acquiring more than one stereo-pair of the scene can be fused into a coherent surface model either directly or via volumetric techniques such as *marching cubes*. In Chapter 8 we conclude the progression from theory to practice, with a number of case examples of 3D vision applications covering areas such as face and body imaging for clinical, veterinary and creative media applications and also 3D vision as a visual prosthetic. An application based only on image matching is also presented that utilizes motion-induced inter-frame disparities within a cine sequence to synthesize missing or damaged frames, or sets of frames, in digitized historic archive footage.

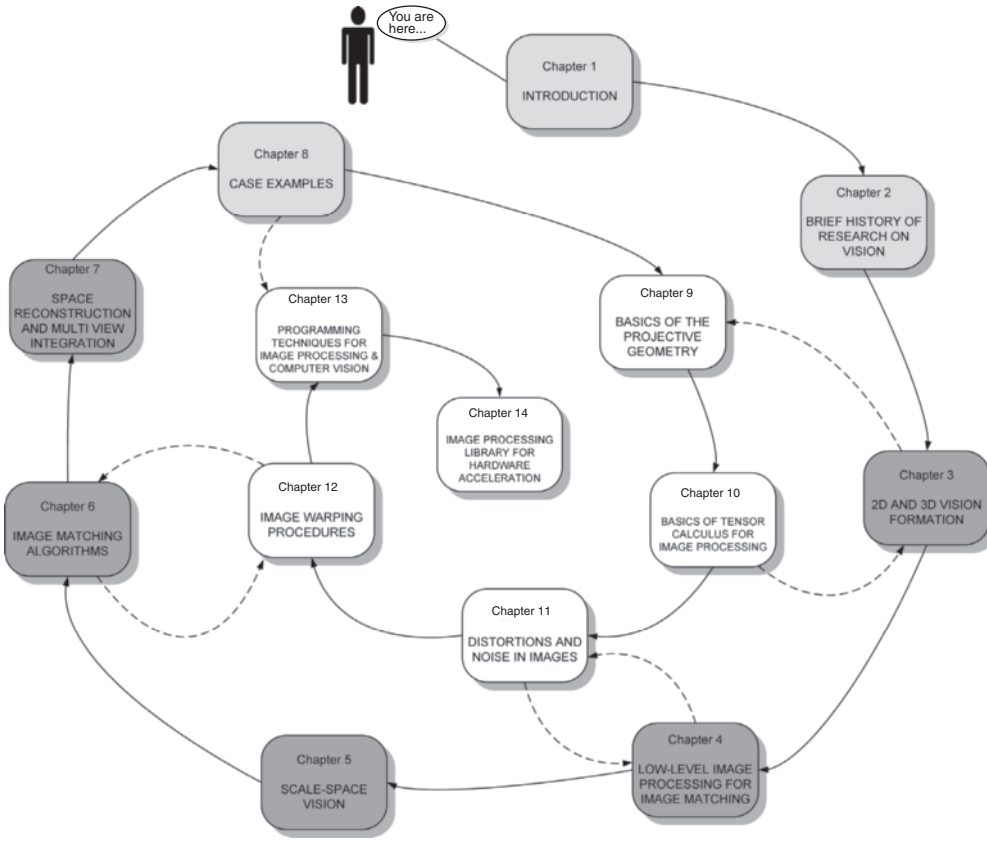

Figure 1.1 Organization of the book

The remaining chapters provide a series of detailed technical tutorials on projective geometry, tensor calculus, image warping procedures and image noise. A chapter on programming techniques for image processing provides practical hints and advice for persons who wish to develop their own computer vision applications. Methods of object oriented programming, such as design patterns, but also proper organization and verification of the code are discussed. Chapter 14 outlines the software presented in the book and provides the link to the recent version of the code.

Figure 1.1 depicts possible order of reading the book. All chapters can be read in number order or selectively as references to specific topics. There are five main chapters (Chapters 3–7), three auxiliary chapters (Chapters 1, 2 and 8) as well as five technical tutorials (Chapters 9–13). The latter are intended to aid understanding of specific topics and can be read in conjunction with the related main chapters, as indicated by the dashed lines in Figure 1.1.