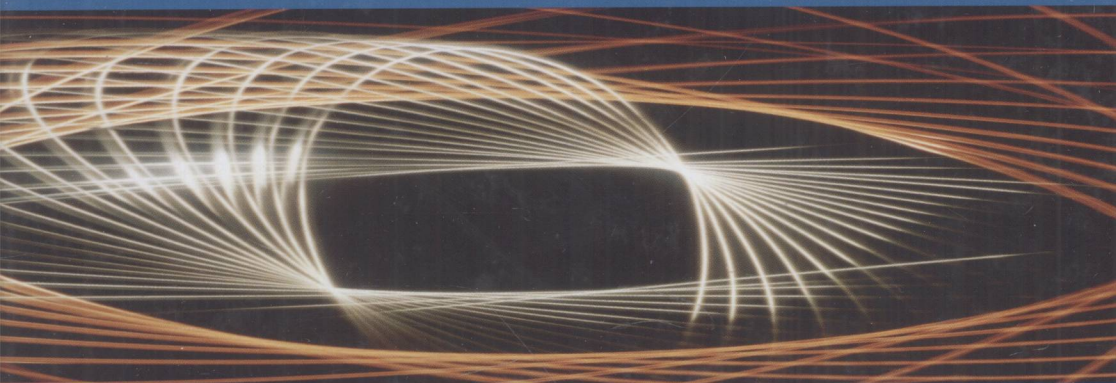


DSP

DIGITAL SIGNAL AND IMAGE PROCESSING SERIES



Visual Perception Through Video Imagery

Edited by Michel Dhome

ISTE

 **WILEY**

Visual Perception through Video Imagery

TN911.7
P428

Visual Perception through Video Imagery

Edited by
Michel Dhome



ISTE



E2009001086

 **WILEY**

First published in France in 2003 by Hermes Science/Lavoisier entitled *Perception visuelle par imagerie video* © LAVOISIER, 2003

First published in Great Britain and the United States in 2009 by ISTE Ltd and John Wiley & Sons, Inc.

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act 1988, this publication may only be reproduced, stored or transmitted, in any form or by any means, with the prior permission in writing of the publishers, or in the case of reprographic reproduction in accordance with the terms and licenses issued by the CLA. Enquiries concerning reproduction outside these terms should be sent to the publishers at the undermentioned address:

ISTE Ltd
27-37 St George's Road
London SW19 4EU
UK
www.iste.co.uk

John Wiley & Sons, Inc.
111 River Street
Hoboken, NJ 07030
USA
www.wiley.com

© ISTE Ltd, 2009

The rights of Michel Dhôme to be identified as the author of this work have been asserted by him in accordance with the Copyright, Designs and Patents Act 1988.

Library of Congress Cataloging-in-Publication Data

[Perception visuelle par imagerie video. English] Visual perception through video imagery / edited by Michel Dhôme.

p. cm.

Includes index.

ISBN 978-1-84821-016-5

1. Computer vision. 2. Visual perception. 3. Vision. I. Dhôme, Michel.

TA1634.P4513 2007

006.3'7--dc22

2007028789

British Library Cataloguing-in-Publication Data

A CIP record for this book is available from the British Library

ISBN: 978-1-84821-016-5

Printed and bound in Great Britain by CPI Antony Rowe, Chippenham and Eastbourne.



Table of Contents

Introduction	13
Part 1	17
Chapter 1. Calibration of Vision Sensors	19
Jean-Marc LAVEST and Gérard RIVES	
1.1. Introduction	19
1.2. General formulation of the problem of calibration	20
1.2.1. Formulation of the problem	20
1.2.1.1. Modeling the camera and lens: pin-hole model	22
1.2.1.2. Formation of images: perspective projection	22
1.2.1.3. Changing lens/camera reference point	23
1.2.1.4. Changing of the camera/image point	24
1.2.1.5. Changing of coordinates in the image plane	24
1.2.2. General expression	25
1.2.2.1. General formulation of the problem of calibration	27
1.3. Linear approach	27
1.3.1. Principle	27
1.3.2. Notes and comments	29
1.4. Non-linear photogrammetric approach	30
1.4.1. Mathematic model	31
1.4.2. Solving the problem	34
1.4.3. Multi-image calibration	35
1.4.4. Self-calibration by bundle adjustment	36
1.4.4.1. Redefinition of the problem	36
1.4.4.2. Estimation of redundancy	37
1.4.4.3. Solution for a near scale factor	37
1.4.4.4. Initial conditions	38

1.4.5. Precision calculation	38
1.5. Results of experimentation	39
1.5.1. Bundle adjustment for a traditional lens	39
1.5.1.1. Initial and experimental conditions	39
1.5.1.2. Sequence of classic images	40
1.5.2. Specific case of fish-eye lenses	42
1.5.2.1. Traditional criterion	43
1.5.2.2. Zero distortion at r_0	43
1.5.2.3. Normalization of distortion coefficients	44
1.5.2.4. Experiments	45
1.5.3. Calibration of underwater cameras	48
1.5.3.1. Theoretical notes	48
1.5.3.2. Experiments	49
1.5.3.3. The material	49
1.5.3.4. Results in air	49
1.5.3.5. Calibration in water	50
1.5.3.6. Relation between the calibration in air and in water . .	53
1.5.4. Calibration of zooms	55
1.5.4.1. Recalling optical properties	55
1.5.4.2. Estimate of the principal point	56
1.5.4.3. Experiments	57
1.6. Bibliography	58
Chapter 2. Self-Calibration of Video Sensors	61
Rachid DERICHE	
2.1. Introduction	61
2.2. Reminder and notation	64
2.3. Huang-Faugeras constraints and Trivedi's equations	66
2.3.1. Huang-Faugeras constraints	66
2.3.2. Trivedi's constraints	67
2.3.3. Discussion	68
2.4. Kruppa equations	68
2.4.1. Geometric derivation of Kruppa equations	68
2.4.2. An algebraic derivation of Kruppa equations	70
2.4.3. Simplified Kruppa equations	72
2.5. Implementation	74
2.5.1. The choice of initial conditions	74
2.5.2. Optimization	75
2.6. Experimental results	76
2.6.1. Estimation of angles and length ratios from images	77

2.6.2. Experiments with synthetic data	78
2.6.3. Experiments with real data	79
2.7. Conclusion	85
2.8. Acknowledgement	87
2.9. Bibliography	87
Chapter 3. Specific Displacements for Self-calibration	91
Diane LINGRAND, François GASPARD and Thierry VIÉVILLE	
3.1. Introduction: interest to resort to specific movements	91
3.2. Modeling: parametrization of specific models	93
3.2.1. Specific projection models	93
3.2.2. Specifications of internal parameters of the camera	96
3.2.3. Taking into account specific displacements	97
3.2.4. Relation with specific properties in the scene	100
3.3. Self-calibration of a camera	100
3.3.1. Usage of pure rotations or points at the horizon	103
3.3.2. Pure rotation and fixed parameters	104
3.3.3. Rotation around a fixed axis	106
3.4. Perception of depth	108
3.4.1. Usage of pure translations	108
3.4.2. Retinal movements	111
3.4.3. Variation of the focal length	114
3.5. Estimating a specific model on real data	119
3.5.1. Application of the estimation mechanism to model inference	122
3.5.2. Some experimental results	123
3.5.3. Application at the localization of a plane	125
3.5.3.1. Rotation in pitch and calibration from a plane	130
3.6. Conclusion	136
3.7. Bibliography	136
Part 2	143
Chapter 4. Localization Tools	145
Michel DHOME and Jean-Thierry LAPRESTÉ	
4.1. Introduction	145
4.2. Geometric modeling of a video camera	146
4.2.1. Pinhole model	146
4.2.2. Perspective projection of a 3D point	147
4.3. Localization of a voluminous object by monocular vision	148
4.3.1. Introduction	148
4.3.2. Mappings	149

4.3.2.1. Matching of lines	149
4.3.2.2. Pairing of points	150
4.3.3. Criterion to minimize	152
4.3.4. Solving the problem using the Newton-Raphson method . .	153
4.3.5. Calculation of partial derivatives	154
4.3.6. Results	156
4.4. Localization of a voluminous object by multi-ocular vision . . .	158
4.4.1. Mathematical developments	158
4.4.2. Calculation of partial derivatives	159
4.4.3. Results	159
4.5. Localization of an articulated object	161
4.5.1. Mathematical development	161
4.5.2. Calculation of partial derivatives for intrinsic parameters . .	163
4.5.3. Results	163
4.6. Hand-eye calibration	164
4.6.1. Introduction	164
4.6.2. Presentation of the method	164
4.6.3. Geometric constraint	166
4.6.4. Results	166
4.7. Initialization methods	168
4.7.1. Initial hypotheses	168
4.7.2. Objective	169
4.7.3. Under the hypothesis of perspective projection	170
4.7.4. Under the hypothesis of scaled orthographic projection . . .	172
4.7.5. Development of the algorithm	173
4.7.6. Specific case of a planar object	174
4.8. Analytical calculations of localization errors	177
4.8.1. Uncertainties in the estimation of a line equation	177
4.8.2. Errors in normals	179
4.8.3. Uncertainties in final localization of polyhedral objects . . .	181
4.8.3.1. Covariance matrix associated with the localization parameters	181
4.9. Conclusion	183
4.10. Bibliography	183
Part 3	187
Chapter 5. Reconstruction of 3D Scenes from Multiple Views	189
Long QUAN, Luce MORIN and Lionel OISEL	
5.1. Introduction	189
5.2. Geometry relating to the acquisition of multiple images	189

5.2.1. Geometry of two images	189
5.2.1.1. Geometric aspect	190
5.2.1.2. Algebraic aspect	191
5.2.1.3. Properties of $\mathbf{F}$	191
5.2.1.4. Estimation of the fundamental matrix	192
5.2.1.5. 7 point algorithm	192
5.2.1.6. 8 point algorithm	193
5.2.1.7. Optimal algorithms	193
5.2.1.8. Robust algorithms which make it possible to eliminate false pairing between a couple of points	194
5.2.2. Geometry of 3 images	195
5.2.3. Geometry beyond 3 images	199
5.3. Matching	200
5.3.1. State of the art elements	200
5.3.1.1. Correlation	201
5.3.1.2. Block-matching	202
5.3.1.3. Dynamic programming	202
5.3.1.4. Association of the optical flow and epipolar geometry	202
5.3.1.5. Energy modeling	204
5.3.2. Dense estimation algorithm based on optical flow	205
5.3.2.1. Hypothesis for the conservation of brightness	205
5.3.2.2. Energy modeling	206
5.3.2.3. Multi-resolution minimization diagram	207
5.4. 3D reconstruction	208
5.4.1. Reconstruction principle: retro-projection	209
5.4.2. Projective reconstruction	209
5.4.3. Euclidean reconstruction	212
5.4.3.1. Calibrated cameras	212
5.4.3.2. Known intrinsic parameters	212
5.4.3.3. Known metric data in the scene	213
5.5. 3D modeling	214
5.5.1. Implicit model	214
5.5.2. Point sets	216
5.5.3. Triangular mesh	216
5.5.3.1. Interactive designation of mesh vertices	217
5.5.3.2. Microfacets	217
5.5.3.3. Triangulation of the points of interest	217
5.5.3.4. Adaptive triangulation	217
5.5.3.5. Regular triangulation	219
5.6. Examples of applications	219

5.6.1. Virtual view rendering	219
5.6.2. VRML models	220
5.7. Conclusion	220
5.8. Bibliography	221
Chapter 6. 3D Reconstruction by Active Dynamic Vision	225
Éric MARCHAND and François CHAUMETTE	
6.1. Introduction: active vision	225
6.2. Reconstruction of 3D primitives	227
6.2.1. Reconstruction by dynamic vision: a rapid state of the art	227
6.2.2. General principle	230
6.2.3. Some specific cases	232
6.2.3.1. Point	232
6.2.3.2. Line	233
6.2.3.3. Cylinder	235
6.2.4. 3D reconstruction by active vision	235
6.2.4.1. 3D reconstruction by active vision: state of the art	236
6.2.4.2. Optimal 3D reconstruction of a primitive	237
6.2.5. Generation of camera movements	240
6.3. Reconstruction of a complete scene	243
6.3.1. Automatic positioning of the camera for the observation of the scene	243
6.3.2. Scene reconstruction: general principle	244
6.3.3. Local focusing strategy	245
6.3.4. Completeness of reconstruction: selection of viewpoints	247
6.3.4.1. Calculation of new viewpoints	247
6.3.4.2. Optimization	250
6.4. Results	250
6.4.1. Reconstruction of 3D primitive: case of the cylinder	251
6.4.2. Perception strategies	252
6.4.2.1. Local exploration	252
6.4.2.2. Total exploration	254
6.5. Conclusion	257
6.6. Appendix: calculation of the interaction matrix	258
6.7. Bibliography	259
Part 4	263
Chapter 7. Shape Recognition in Images	265
Patrick GROS and Cordelia SCHMID	
7.1. Introduction	265

7.2. State of the art	266
7.2.1. Searching images based on photometric data	266
7.2.2. Search for images based on geometric data	267
7.2.3. Recognition using a 3D geometric model	268
7.2.4. Recognition using a set of images	270
7.3. Principle of local quasi-invariants	270
7.4. Photometric approach	272
7.4.1. Key points	272
7.4.2. Differential invariants of gray levels	273
7.4.3. Comparison of descriptors with Mahalanobis distance	275
7.4.4. Voting algorithm	276
7.4.5. Semi-local constraints	277
7.4.6. Multi-dimensional indexing	278
7.4.7. Experimental results	279
7.4.8. Extensions	282
7.5. Geometric approach	284
7.5.1. Basic algorithm	284
7.5.2. Some results	285
7.5.2.1. Pairing results	285
7.5.2.2. Results of indexing and recognition	286
7.6. Indexing of images	288
7.6.1. Traditional approaches	290
7.6.2. VA-File and the Pyramid-Tree	291
7.6.3. Some results	292
7.6.3.1. Context of experiments	293
7.6.3.2. First experiment	293
7.6.3.3. Second experiment	293
7.6.3.4. Third experiment	294
7.6.4. Some prospects	294
7.7. Conclusion	295
7.8. Bibliography	296
List of Authors	301
Index	305

Introduction

Artificial vision with a main objective of automatic perception and interpretation of the universe observed by a system containing one or several cameras is a relatively new field of investigation. It leads to a surprisingly large range of problems, and most of these are not currently resolved in a reliable way. Although a general theory is not close to being reached, significant progress has been made recently, theoretically as well as methodologically.

In the visible world, images of luminance are the result of two physical processes: the first one is linked to reflectance properties of the surface of observed objects, while the second one corresponds to the projection of these same objects on the light sensitive plate of the sensor used. From a mathematical standpoint, in order to interpret the observed scene, we must solve an inverse problem, i.e. infer the surface geometry (3D) of objects present, from the purely 2D content of the image or from logged images.

This reputedly complex problem in the context of computer vision is solved by man with surprising ease. However, the human vision system operation is clearly not founded on a single concept. Examining the implemented processes during short or long distance vision is sufficient proof. In the first case, the existing disparity between left and right retinal images makes it possible for man to obtain indepth information by triangulation (*stereoscopy*) relating to its close environment which is vital in particular to manually capture objects. In the second case, when looking at long distance, or even more so when contemplating a picture, stereoscopy is obviously no help in interpreting the observed scene. Even under these conditions, however (*total lack of direct 3D information*), man is able to estimate the form and spatial position of objects he observes in the vast majority of cases. This requires mental processes, from 2D

information extracted from a luminance image, able to infer 3D information. These are based on the unconscious use of prior knowledge relating to the principle of retinal image composition and the form of 3D objects surrounding us. The surprising capabilities of the human vision system are because this knowledge is continuously enhanced from early childhood.

In the last few decades, researchers in the artificial vision community have attempted to develop perception systems that would work from data emanating from video cameras. This book presents a few tools emerging from recent advances in the field.

In Part 1, the reader will find three chapters dedicated to *calibration* or *self-calibration* of video sensors. Chapter 1 presents a finite estimation approach of the intrinsic parameters of a video camera, greatly inspired by the world of photogrammetry. It is based on the interpretation of images from a calibration test chart which is not generally known with great precision. Chapter 2 addresses the complex self-calibration problem from a series of matching points between different images from a single scene. The recommended method is based on an elegant and simple decomposition of Kruppa equations. Chapter 3 explores the self-calibration problem of cameras with *specific movements* making the implementation of simplified development of the main matrix possible.

Part 2 mainly involves the estimation of the *relative object/sensor position* by introducing prior knowledge (CAD object model). The reader will discover how to treat the localization problem of a rigid object observed by a monocular system. The formalism presented is then extended to understand cases as different as multi-ocular localization and hand-eye calibration, and research on the posture of articulated objects such as robotic arms.

Part 3 addresses volume information inference in two chapters. Chapter 5 discusses the *reconstruction* problem of a fixed scene observed by a multi-ocular system. The notions homo-log points, epipolar geometry, fundamental matrix, essential matrix and trifocal tensor are first introduced as well as different approaches for obtaining these entities. The problem of dense matching between image pairs is addressed before a few reconstruction examples are shown. Chapter 6 discusses the notion of *active dynamic vision*. It is the control of the path of a camera embedded in a robotic arm in order to reconstruct the surrounding scene. An underlying problem involves the definition of optimal movements necessary for the reconstruction of different primitives (points, straight lines, cylinders). Finally, perception strategies are

proposed in order to ensure a complete reconstruction taking inter-object blanking into consideration.

The *recognition of forms in images* is the heart of Part 4. Chapter 7 is dedicated to proposing tools for the identification in a database of the images containing visual elements identical to those contained in a request image; the differences of acquisition between images which could involve the point of view, conditions of illumination and global composition of the scene observed. The methods presented are based on a common principle: the use of quasi-invariants associated with local descriptors. These are tested against large image databases.

