

SECOND EDITION

Multiple View Geometry

in computer vision



Richard Hartley and Andrew Zisserman

CAMBRIDGE

Multiple View Geometry in Computer Vision

Second Edition

Richard Hartley
Australian National University,
Canberra, Australia

Andrew Zisserman
University of Oxford, UK

江苏工业学院图书馆
藏书章

 CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge, CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa
<http://www.cambridge.org>

© Cambridge University Press 2000, 2003

This book is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without the
written permission of Cambridge University Press.

First Published 2000
Reprinted 2001, 2002
Second Edition 2003

Printed in the United Kingdom at the University Press, Cambridge

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

Library of Congress Cataloguing in Publication data

ISBN 0521 54051 8 hardback

Dedication

This book is dedicated to Joe Mundy whose vision and constant search for new ideas led us into this field.

Foreword

By Olivier Faugeras

Making a computer see was something that leading experts in the field of Artificial Intelligence thought to be at the level of difficulty of a summer student's project back in the sixties. Forty years later the task is still unsolved and seems formidable. A whole field, called Computer Vision, has emerged as a discipline in itself with strong connections to mathematics and computer science and looser connections to physics, the psychology of perception and the neuro sciences.

One of the likely reasons for this half-failure is the fact that researchers had overlooked the fact, perhaps because of this plague called naive introspection, that perception in general and visual perception in particular are far more complex in animals and humans than was initially thought. There is of course no reason why we should pattern Computer Vision algorithms after biological ones, but the fact of the matter is that

- (i) the way biological vision works is still largely unknown and therefore hard to emulate on computers, and
- (ii) attempts to ignore biological vision and reinvent a sort of silicon-based vision have not been so successful as initially expected.

Despite these negative remarks, Computer Vision researchers have obtained some outstanding successes, both practical and theoretical.

On the side of practice, and to single out one example, the possibility of guiding vehicles such as cars and trucks on regular roads or on rough terrain using computer vision technology was demonstrated many years ago in Europe, the USA and Japan. This requires capabilities for real-time three-dimensional dynamic scene analysis which are quite elaborate. Today, car manufacturers are slowly incorporating some of these functions in their products.

On the theoretical side some remarkable progress has been achieved in the area of what one could call geometric Computer Vision. This includes the description of the way the appearance of objects changes when viewed from different viewpoints as a function of the objects' shape and the cameras parameters. This endeavour would not have been achieved without the use of fairly sophisticated mathematical techniques encompassing many areas of geometry, ancient and novel. This book deals in particular with the intricate and beautiful geometric relations that exist between the images of objects in the world. These relations are important to analyze for their own sake because

this is one of the goals of science to provide explanations for appearances; they are also important to analyze because of the range of applications their understanding opens up.

The book has been written by two pioneers and leading experts in geometric Computer Vision. They have succeeded in what was something of a challenge, namely to convey in a simple and easily accessible way the mathematics that is necessary for understanding the underlying geometric concepts, to be quite exhaustive in the coverage of the results that have been obtained by them and other researchers worldwide, to analyze the interplay between the geometry and the fact that the image measurements are necessarily noisy, to express many of these theoretical results in algorithmic form so that they can readily be transformed into computer code, and to present many real examples that illustrate the concepts and show the range of applicability of the theory.

Returning to the original holy grail of making a computer see we may wonder whether this kind of work is a step in the right direction. I must leave the readers of the book to answer this question, and be content with saying that no designer of systems using cameras hooked to computers that will be built in the foreseeable future can ignore this work. This is perhaps a step in the direction of defining what it means for a computer to see.

Preface

Over the past decade there has been a rapid development in the understanding and modelling of the geometry of multiple views in computer vision. The theory and practice have now reached a level of maturity where excellent results can be achieved for problems that were certainly unsolved a decade ago, and often thought unsolvable. These tasks and algorithms include:

- Given two images, and no other information, compute matches between the images, and the 3D position of the points that generate these matches and the cameras that generate the images.
- Given three images, and no other information, similarly compute the matches between images of points and lines, and the position in 3D of these points and lines and the cameras.
- Compute the epipolar geometry of a stereo rig, and trifocal geometry of a trinocular rig, without requiring a calibration object.
- Compute the internal calibration of a camera from a sequence of images of natural scenes (i.e. calibration “on the fly”).

The distinctive flavour of these algorithms is that they are *uncalibrated* — it is not necessary to know or first need to compute the camera internal parameters (such as the focal length).

Underpinning these algorithms is a new and more complete theoretical understanding of the geometry of multiple uncalibrated views: the number of parameters involved, the constraints between points and lines imaged in the views; and the retrieval of cameras and 3-space points from image correspondences. For example, to determine the epipolar geometry of a stereo rig requires specifying only seven parameters, the camera calibration is not required. These parameters are determined from the correspondence of seven or more image point correspondences. Contrast this uncalibrated route, with the previous calibrated route of a decade ago: each camera would first be calibrated from the image of a carefully engineered calibration object with known geometry. The calibration involves determining 11 parameters for each camera. The epipolar geometry would then have been computed from these two sets of 11 parameters.

This example illustrates the importance of the uncalibrated (projective) approach — using the appropriate representation of the geometry makes explicit the parameters

that are required at each stage of a computation. This avoids computing parameters that have no effect on the final result, and results in simpler algorithms. It is also worth correcting a possible misconception. In the uncalibrated framework, entities (for instance point positions in 3-space) are often recovered to within a precisely defined ambiguity. This ambiguity does not mean that the points are poorly estimated.

More practically, it is often not possible to calibrate cameras once-and-for-all; for instance where cameras are moved (on a mobile vehicle) or internal parameters are changed (a surveillance camera with zoom). Furthermore, calibration information is simply not available in some circumstances. Imagine computing the motion of a camera from a video sequence, or building a virtual reality model from archive film footage where both motion and internal calibration information are unknown.

The achievements in multiple view geometry have been possible because of developments in our theoretical understanding, but also because of improvements in estimating mathematical objects from images. The first improvement has been an attention to the error that should be minimized in over-determined systems – whether it be algebraic, geometric or statistical. The second improvement has been the use of robust estimation algorithms (such as RANSAC), so that the estimate is unaffected by “outliers” in the data. Also these techniques have generated powerful search and matching algorithms.

Many of the problems of reconstruction have now reached a level where we may claim that they are solved. Such problems include:

- (i) Estimation of the multifocal tensors from image point correspondences, particularly the fundamental matrix and trifocal tensors (the quadrifocal tensor having not received so much attention).
- (ii) Extraction of the camera matrices from these tensors, and subsequent projective reconstruction from two, three and four views.

Other significant successes have been achieved, though there may be more to learn about these problems. Examples include:

- (i) Application of bundle adjustment to solve more general reconstruction problems.
- (ii) Metric (Euclidean) reconstruction given minimal assumptions on the camera matrices.
- (iii) Automatic detection of correspondences in image sequences, and elimination of outliers and false matches using the multifocal tensor relationships.

Roadplan. The book is divided into six parts and there are seven short appendices. Each part introduces a new geometric relation: the homography for background, the camera matrix for single view, the fundamental matrix for two views, the trifocal tensor for three views, and the quadrifocal tensor for four views. In each case there is a chapter describing the relation, its properties and applications, and a companion chapter describing algorithms for its estimation from image measurements. The estimation algorithms described range from cheap, simple, approaches through to the optimal algorithms which are currently believed to be the best available.

Part 0: Background. This part is more tutorial than the others. It introduces the central ideas in the projective geometry of 2-space and 3-space (for example ideal points, and the absolute conic); how this geometry may be represented, manipulated, and estimated; and how the geometry relates to various objectives in computer vision such as rectifying images of planes to remove perspective distortion.

Part 1: Single view geometry. Here the various cameras that model the perspective projection from 3-space to an image are defined and their anatomy explored. Their estimation using traditional techniques of calibration objects is described, as well as camera calibration from vanishing points and vanishing lines.

Part 2: Two view geometry. This part describes the epipolar geometry of two cameras, projective reconstruction from image point correspondences, methods of resolving the projective ambiguity, optimal triangulation, transfer between views via planes.

Part 3: Three view geometry. Here the trifocal geometry of three cameras is described, including transfer of a point correspondence from two views to a third, and similarly transfer for a line correspondence; computation of the geometry from point and line correspondences, retrieval of the camera matrices.

Part 4: N-views. This part has two purposes. First, it extends three view geometry to four views (a minor extension) and describes estimation methods applicable to N-views, such as the factorization algorithm of Tomasi and Kanade for computing structure and motion simultaneously from multiple images. Second, it covers themes that have been touched on in earlier chapters, but can be understood more fully and uniformly by emphasising their commonality. Examples include deriving multi-linear view constraints on correspondences, auto-calibration, and ambiguous solutions.

Appendices. These describe further background material on tensors, statistics, parameter estimation, linear and matrix algebra, iterative estimation, the solution of sparse matrix systems, and special projective transformations.

Acknowledgements. We have benefited enormously from ideas and discussions with our colleagues: Paul Beardsley, Stefan Carlsson, Olivier Faugeras, Andrew Fitzgibbon, Jitendra Malik, Steve Maybank, Amnon Shashua, Phil Torr, Bill Triggs.

If there are only a countable number of errors in this book then it is due to Antonio Criminisi, David Liebowitz and Frederik Schaffalitzky who have with great energy and devotion read most of it, and made numerous suggestions for improvements. Similarly both Peter Sturm and Bill Triggs have suggested many improvements to various chapters. We are grateful to other colleagues who have read individual chapters: David Capel, Lourdes de Agapito Vicente, Bob Kaucic, Steve Maybank, Peter Tu.

We are particularly grateful to those who have provided multiple figures: Paul Beardsley, Antonio Criminisi, Andrew Fitzgibbon, David Liebowitz, and Larry Shapiro; and for individual figures from: Martin Armstrong, David Capel, Lourdes de Agapito Vicente, Eric Hayman, Phil Pritchett, Luc Robert, Cordelia Schmid, and others who are explicitly acknowledged in figure captions.

At Cambridge University Press we thank David Tranah for his constant source of advice and patience, and Michael Behrend for excellent copy editing.

A small number of minor errors have been corrected in the reprinted editions, and we thank the following readers for pointing these out: Luis Baumela, Niclas Borlin, Mike Brooks, Jun ho. Choi, Wojciech Chojnacki, Carlo Colombo, Nicolas Dano, Andrew Fitzgibbon, Bogdan Georgescu, Fredrik Kahl, Bob Kaucic, Jae-Hak Kim, Han-sung Lee, Dennis Maier, Karsten Muelhmann, David Nister, Andreas Olsson, Stéphane Paris, Frederik Schaffalitzky, Bill Severson, Pedro Lopez de Teruel Alcolea, Bernard Thiesse, Ken Thornton, Magdalena Urbanek, Gergely Vass, Eugene Vendrovsky, Sui Wei, and Tomáš Werner.

The second edition. This new paperback edition has been expanded to include some of the developments since the original version of July 2000. For example, the book now covers the discovery of a closed form factorization solution in the projective case when a plane is visible in the scene, and the extension of affine factorization to non-rigid scenes. We have also extended the discussion of single view geometry (chapter 8) and three view geometry (chapter 15), and added an appendix on parameter estimation.

In preparing this second edition we are very grateful to colleagues who have made suggestion for improvements and additions. These include Marc Pollefeys, Bill Triggs and in particular Tomáš Werner who provided excellent and comprehensive comments. We also thank Antonio Criminisi, Andrew Fitzgibbon, Rob Fergus, David Liebowitz, and particularly Josef Šivic, for proof reading and very helpful comments on parts of the new material. As always we are grateful to David Tranah of CUP.

The figures appearing in this book can be downloaded from

<http://www.robots.ox.ac.uk/~vgg/hzbook.html>

This site also includes Matlab code for several of the algorithms, and lists the errata of earlier printings.

I am never forget the day my first book is published. Every chapter I stole from somewhere else. Index I copy from old Vladivostok telephone directory. This book, this book was sensational!

Excerpts from “Nikolai Ivanovich Lobachevsky” by Tom Lehrer.

Contents

<i>Foreword</i>	<i>page</i> xi
<i>Preface</i>	xiii
1 Introduction – a Tour of Multiple View Geometry	1
1.1 Introduction – the ubiquitous projective geometry	1
1.2 Camera projections	6
1.3 Reconstruction from more than one view	10
1.4 Three-view geometry	12
1.5 Four view geometry and n -view reconstruction	13
1.6 Transfer	14
1.7 Euclidean reconstruction	16
1.8 Auto-calibration	17
1.9 The reward I : 3D graphical models	18
1.10 The reward II: video augmentation	19
PART 0: The Background: Projective Geometry, Transformations and Esti- mation	23
<i>Outline</i>	24
2 Projective Geometry and Transformations of 2D	25
2.1 Planar geometry	25
2.2 The 2D projective plane	26
2.3 Projective transformations	32
2.4 A hierarchy of transformations	37
2.5 The projective geometry of 1D	44
2.6 Topology of the projective plane	46
2.7 Recovery of affine and metric properties from images	47
2.8 More properties of conics	58
2.9 Fixed points and lines	61
2.10 Closure	62
3 Projective Geometry and Transformations of 3D	65
3.1 Points and projective transformations	65
3.2 Representing and transforming planes, lines and quadrics	66

3.3	Twisted cubics	75
3.4	The hierarchy of transformations	77
3.5	The plane at infinity	79
3.6	The absolute conic	81
3.7	The absolute dual quadric	83
3.8	Closure	85
4	Estimation – 2D Projective Transformations	87
4.1	The Direct Linear Transformation (DLT) algorithm	88
4.2	Different cost functions	93
4.3	Statistical cost functions and Maximum Likelihood estimation	102
4.4	Transformation invariance and normalization	104
4.5	Iterative minimization methods	110
4.6	Experimental comparison of the algorithms	115
4.7	Robust estimation	116
4.8	Automatic computation of a homography	123
4.9	Closure	127
5	Algorithm Evaluation and Error Analysis	132
5.1	Bounds on performance	132
5.2	Covariance of the estimated transformation	138
5.3	Monte Carlo estimation of covariance	149
5.4	Closure	150
	PART I: Camera Geometry and Single View Geometry	151
	<i>Outline</i>	152
6	Camera Models	153
6.1	Finite cameras	153
6.2	The projective camera	158
6.3	Cameras at infinity	166
6.4	Other camera models	174
6.5	Closure	176
7	Computation of the Camera Matrix P	178
7.1	Basic equations	178
7.2	Geometric error	180
7.3	Restricted camera estimation	184
7.4	Radial distortion	189
7.5	Closure	193
8	More Single View Geometry	195
8.1	Action of a projective camera on planes, lines, and conics	195
8.2	Images of smooth surfaces	200
8.3	Action of a projective camera on quadrics	201
8.4	The importance of the camera centre	202
8.5	Camera calibration and the image of the absolute conic	208

8.6	Vanishing points and vanishing lines	213
8.7	Affine 3D measurements and reconstruction	220
8.8	Determining camera calibration K from a single view	223
8.9	Single view reconstruction	229
8.10	The calibrating conic	231
8.11	Closure	233
PART II: Two-View Geometry		237
	<i>Outline</i>	238
9	Epipolar Geometry and the Fundamental Matrix	239
9.1	Epipolar geometry	239
9.2	The fundamental matrix F	241
9.3	Fundamental matrices arising from special motions	247
9.4	Geometric representation of the fundamental matrix	250
9.5	Retrieving the camera matrices	253
9.6	The essential matrix	257
9.7	Closure	259
10	3D Reconstruction of Cameras and Structure	262
10.1	Outline of reconstruction method	262
10.2	Reconstruction ambiguity	264
10.3	The projective reconstruction theorem	266
10.4	Stratified reconstruction	267
10.5	Direct reconstruction – using ground truth	275
10.6	Closure	276
11	Computation of the Fundamental Matrix F	279
11.1	Basic equations	279
11.2	The normalized 8-point algorithm	281
11.3	The algebraic minimization algorithm	282
11.4	Geometric distance	284
11.5	Experimental evaluation of the algorithms	288
11.6	Automatic computation of F	290
11.7	Special cases of F -computation	293
11.8	Correspondence of other entities	294
11.9	Degeneracies	295
11.10	A geometric interpretation of F -computation	297
11.11	The envelope of epipolar lines	298
11.12	Image rectification	302
11.13	Closure	308
12	Structure Computation	310
12.1	Problem statement	310
12.2	Linear triangulation methods	312
12.3	Geometric error cost function	313
12.4	Sampson approximation (first-order geometric correction)	314

12.5	An optimal solution	315
12.6	Probability distribution of the estimated 3D point	321
12.7	Line reconstruction	321
12.8	Closure	323
13	Scene planes and homographies	325
13.1	Homographies given the plane and vice versa	326
13.2	Plane induced homographies given F and image correspondences	329
13.3	Computing F given the homography induced by a plane	334
13.4	The infinite homography H_∞	338
13.5	Closure	340
14	Affine Epipolar Geometry	344
14.1	Affine epipolar geometry	344
14.2	The affine fundamental matrix	345
14.3	Estimating F_A from image point correspondences	347
14.4	Triangulation	353
14.5	Affine reconstruction	353
14.6	Necker reversal and the bas-relief ambiguity	355
14.7	Computing the motion	357
14.8	Closure	360
PART III:	Three-View Geometry	363
	<i>Outline</i>	364
15	The Trifocal Tensor	365
15.1	The geometric basis for the trifocal tensor	365
15.2	The trifocal tensor and tensor notation	376
15.3	Transfer	379
15.4	The fundamental matrices for three views	383
15.5	Closure	387
16	Computation of the Trifocal Tensor \mathcal{T}	391
16.1	Basic equations	391
16.2	The normalized linear algorithm	393
16.3	The algebraic minimization algorithm	395
16.4	Geometric distance	396
16.5	Experimental evaluation of the algorithms	399
16.6	Automatic computation of \mathcal{T}	400
16.7	Special cases of \mathcal{T} -computation	404
16.8	Closure	406
PART IV:	N-View Geometry	409
	<i>Outline</i>	410
17	N-Linearities and Multiple View Tensors	411
17.1	Bilinear relations	411
17.2	Trilinear relations	414

17.3	Quadrilinear relations	418
17.4	Intersections of four planes	421
17.5	Counting arguments	422
17.6	Number of independent equations	428
17.7	Choosing equations	431
17.8	Closure	432
18	<i>N</i>-View Computational Methods	434
18.1	Projective reconstruction – bundle adjustment	434
18.2	Affine reconstruction – the factorization algorithm	436
18.3	Non-rigid factorization	440
18.4	Projective factorization	444
18.5	Projective reconstruction using planes	447
18.6	Reconstruction from sequences	452
18.7	Closure	456
19	Auto-Calibration	458
19.1	Introduction	458
19.2	Algebraic framework and problem statement	459
19.3	Calibration using the absolute dual quadric	462
19.4	The Kruppa equations	469
19.5	A stratified solution	473
19.6	Calibration from rotating cameras	481
19.7	Auto-calibration from planes	485
19.8	Planar motion	486
19.9	Single axis rotation – turntable motion	490
19.10	Auto-calibration of a stereo rig	493
19.11	Closure	497
20	Duality	502
20.1	Carlsson–Weinshall duality	502
20.2	Reduced reconstruction	508
20.3	Closure	513
21	Cheirality	515
21.1	Quasi-affine transformations	515
21.2	Front and back of a camera	518
21.3	Three-dimensional point sets	519
21.4	Obtaining a quasi-affine reconstruction	520
21.5	Effect of transformations on cheirality	521
21.6	Orientation	523
21.7	The cheiral inequalities	525
21.8	Which points are visible in a third view	528
21.9	Which points are in front of which	530
21.10	Closure	531

22 Degenerate Configurations	533
22.1 Camera resectioning	533
22.2 Degeneracies in two views	539
22.3 Carlsson–Weinshall duality	546
22.4 Three-view critical configurations	553
22.5 Closure	558
 PART V : Appendices	561
<i>Appendix 1</i> Tensor Notation	562
<i>Appendix 2</i> Gaussian (Normal) and χ^2 Distributions	565
<i>Appendix 3</i> Parameter Estimation	568
<i>Appendix 4</i> Matrix Properties and Decompositions	578
<i>Appendix 5</i> Least-squares Minimization	588
<i>Appendix 6</i> Iterative Estimation Methods	597
<i>Appendix 7</i> Some Special Plane Projective Transformations	628
<i>Bibliography</i>	634
<i>Index</i>	646

Introduction – a Tour of Multiple View Geometry

This chapter is an introduction to the principal ideas covered in this book. It gives an *informal* treatment of these topics. Precise, unambiguous definitions, careful algebra, and the description of well honed estimation algorithms is postponed until chapter 2 and the following chapters in the book. Throughout this introduction we will generally not give specific forward pointers to these later chapters. The material referred to can be located by use of the index or table of contents.

1.1 Introduction – the ubiquitous projective geometry

We are all familiar with projective transformations. When we look at a picture, we see squares that are not squares, or circles that are not circles. The transformation that maps these planar objects onto the picture is an example of a projective transformation.

So what properties of geometry are preserved by projective transformations? Certainly, shape is not, since a circle may appear as an ellipse. Neither are lengths since two perpendicular radii of a circle are stretched by different amounts by the projective transformation. Angles, distance, ratios of distances – none of these are preserved, and it may appear that very little geometry is preserved by a projective transformation. However, a property that is preserved is that of straightness. It turns out that this is the most general requirement on the mapping, and we may define a projective transformation of a plane as any mapping of the points on the plane that preserves straight lines.

To see why we will require projective geometry we start from the familiar Euclidean geometry. This is the geometry that describes angles and shapes of objects. Euclidean geometry is troublesome in one major respect – we need to keep making an exception to reason about some of the basic concepts of the geometry – such as intersection of lines. Two lines (we are thinking here of 2-dimensional geometry) almost always meet in a point, but there are some pairs of lines that do not do so – those that we call parallel. A common linguistic device for getting around this is to say that parallel lines meet “at infinity”. However this is not altogether convincing, and conflicts with another dictum, that infinity does not exist, and is only a convenient fiction. We can get around this by