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, Hansung 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.