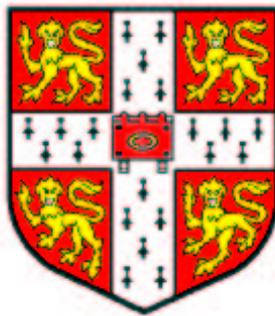


Multiview Geometry: Profiles and Self-Calibration

Paulo Ricardo dos Santos Mendonça

Jesus College

Lent Term 2001



A dissertation submitted to the University of Cambridge
for the degree of Doctor of Philosophy

To Cláudia
with all my love.

Declaration

I hereby declare that no part of this thesis has already been or is being submitted for any other degree or qualification. This dissertation is the result of my own original work carried out in the Department of Engineering of the University of Cambridge, except where explicit reference has been made to the work of others. The dissertation contains 33000 words and 41 figures.

Acknowledgements

This thesis is the result of the research I carried out at the Department of Engineering of the University of Cambridge, under the supervision of Prof. Roberto Cipolla. During this period I have benefited from the help of many people, but most notably Prof. Cipolla, who has been a most supportive and enthusiastic supervisor, and my friend and fellow PhD student Kenneth Wong, with whom the sharing of authorship in so many papers truthfully reflects the closeness of our cooperation. I am earnestly grateful to both of them.

A number of researchers have contributed to this work through discussions and criticisms, including Dr. Ezio Malis, Prof. Andrew Zisserman, Dr. Harpreet Sawhney, Dr. Richard Hartley, Dr. Geoff Cross, Dr. Keith Hanna, Dr. Paul Beardsley, Dr. Joe Mundy and Prof. Andrew Blake. Many people from the Engineering Department have also given their invaluable help, such as Dr. Tom Drummond, Dr. Joan Lasenby and Dr. Yongduek Seo, and my colleagues Adrian Broadhurst, Paul Smith, Anthony Dick and Duncan Robertson. I am indebted to CAPES/Brazilian Ministry of Education, whose financial support made my stay in Cambridge possible, and to Jesus College and the Department of Engineering for providing generous aid to my participation in conferences, workshops and seminars.

I also wish to thank Sam and Robert McLaughlin, Esther and Guido Timm, Monica and Jorge Karadima, Maria and Marco Kampp, and Graziano Chesi for their precious and cheerful friendship.

Finally, my heartfelt gratitude to my parents, Paulo and Ruth, and my brother André, for their immense incentive, and to my wife Cláudia, who makes it all worthwhile.

Paulo R. S. Mendonça.

Contents

| | |
|--|-------------|
| Acknowledgements | vii |
| Abstract | xiii |
| List of Figures | xv |
| Glossary | xix |
| 1 Introduction | 1 |
| 1.1 Motivation | 1 |
| 1.1.1 Shape from X | 1 |
| 1.1.2 Geometry-Based Algorithms | 2 |
| 1.1.3 Structure and Motion from Profiles | 3 |
| 1.1.4 Camera Self-Calibration | 3 |
| 1.2 Approach | 4 |
| 1.2.1 Models and Approximations Adopted | 5 |
| 1.3 Contributions | 6 |
| 1.4 Outline of the Thesis | 7 |
| 2 Projective Geometry | 9 |
| 2.1 Introduction | 9 |

| | | |
|----------|--|-----------|
| 2.2 | Homogeneous Coordinates | 10 |
| 2.2.1 | Lines and Planes in Homogeneous Coordinates | 12 |
| 2.2.2 | Conics and Quadrics in Homogeneous Coordinates | 13 |
| 2.3 | Projective Transformations | 15 |
| 2.3.1 | Projective Cameras | 17 |
| 2.3.2 | Affine Cameras | 20 |
| 2.4 | Projection of Planes and Surfaces | 20 |
| 2.4.1 | Projection of Implicit Surfaces | 21 |
| 2.5 | Summary and Conclusions | 24 |
| 3 | Epipolar Geometry and Multicamera Systems | 27 |
| 3.1 | Introduction | 28 |
| 3.1.1 | The Fundamental Matrix | 28 |
| 3.1.2 | Plane Plus Parallax Representation | 30 |
| 3.1.3 | The Essential Matrix | 33 |
| 3.1.4 | Computation of the Fundamental Matrix | 34 |
| 3.2 | Affine Epipolar Geometry | 38 |
| 3.2.1 | The Affine Fundamental Matrix | 38 |
| 3.3 | Epipolar Geometry and Smooth Surfaces | 41 |
| 3.3.1 | Affine Epipolar Geometry of Smooth Surfaces | 43 |
| 3.4 | Geometry of Multiple Cameras | 45 |
| 3.4.1 | Trifocal Tensors | 46 |
| 3.4.2 | General Multilinear Relations | 52 |
| 3.5 | Conclusions | 53 |
| 4 | Epipolar Geometry from Profiles Under Circular Motion | 55 |

| | | |
|----------|---|-----------|
| 4.1 | Introduction | 55 |
| 4.2 | Previous Works and Bibliographic Review | 57 |
| 4.3 | Symmetry in the Image of a Surface of Revolution | 60 |
| 4.4 | Parameterisations of the Fundamental Matrix | 65 |
| 4.4.1 | Fundamental Matrix under Circular Motion | 65 |
| 4.4.2 | Parameterisation via Planar Harmonic Homology | 68 |
| 4.5 | Motion Estimation | 71 |
| 4.5.1 | Estimation of the Epipoles Constrained to the Horizon | 76 |
| 4.5.2 | Limitations of the Algorithm | 77 |
| 4.6 | Implementation and Experimental Results | 79 |
| 4.7 | Reconstruction from Image Profiles | 82 |
| 4.7.1 | Description of the Method | 83 |
| 4.7.2 | Implementation and Experimental Results | 84 |
| 4.8 | Summary and Conclusions | 85 |
| 5 | Camera Self-Calibration | 89 |
| 5.1 | Introduction | 89 |
| 5.2 | Previous Works | 90 |
| 5.3 | Theoretical Background | 92 |
| 5.3.1 | The Huang and Faugeras Constraints | 92 |
| 5.4 | Self-Calibration from the Essential Matrix | 93 |
| 5.4.1 | Linear Solution | 94 |
| 5.4.2 | Detection of Critical Motions | 97 |
| 5.4.3 | Nonlinear Solution | 102 |
| 5.5 | Description of the Algorithm | 105 |
| 5.6 | Experimental Results | 106 |

| | | |
|----------|---|------------|
| 5.6.1 | Synthetic Data | 106 |
| 5.6.2 | Real Data | 109 |
| 5.7 | Conclusions | 114 |
| 6 | Conclusion | 117 |
| 6.1 | Summary | 117 |
| 6.2 | Future Work | 119 |
| A | Derivatives of the Geometric Error of the Fundamental Matrix | 121 |
| A.1 | Gradient of the Geometric Error | 122 |
| A.2 | Hessian of the Geometric Error | 124 |
| B | Derivative of the Huang and Faugeras Constraints | 127 |
| | Bibliography | 129 |
| | Index | 151 |

Multiview Geometry: Profiles and Self-Calibration

Abstract

An important goal of computer vision is the simultaneous recovery of camera motion and scene structure from multiple views. Frequently, the solution of this problem demands the estimation of the *epipolar geometry* of the sequence of images, encoded in *fundamental matrices*. These can be satisfactorily obtained through a number of methods when image features corresponding to the same objects in space — such as points, lines, texture etc. — can be easily matched. However, when the scene is comprised of smooth, textureless surfaces, the determination of such correspondences is a difficult problem.

In this situation the most prominent features of the objects being viewed are the *profiles* or *apparent contours*. This dissertation develops an efficient technique to estimate the epipolar geometry from profiles in the important case of circular motion. In contrast to previous methods, the solution proposed here can be used even for surfaces with simple geometry. It also employs a sequential approach, obtaining the independent components of the epipolar geometry step-by-step. Finally, it does not demand the solution of any large scale optimisation problem.

Once the epipolar geometry is estimated, the projective structure of the scene can be determined. To update this projective reconstruction to an Euclidean one, the camera must be calibrated. In this dissertation a novel self-calibration technique is introduced, based on obtaining the intrinsic parameters that update the fundamental matrices to *essential matrices*. This simple approach provides an algorithm for the linear computation of the varying focal lengths of the cameras, plus a nonlinear method that can refine the initial solution and also obtain other intrinsic parameters.

List of Figures

| | | |
|-----|--|----|
| 2.1 | Polar of a quadric with respect to a point. | 14 |
| 2.2 | Geometric representation of a projective camera. | 19 |
| 2.3 | 3D image of torus and camera. | 24 |
| 2.4 | Projection of a torus. | 25 |
| 3.1 | Geometric representation of the epipolar constraint. | 30 |
| 3.2 | Geometric error on the computation of the fundamental matrix. | 36 |
| 3.3 | Frontier point, contour generator, profiles and epipolar tangencies. | 42 |
| 3.4 | Geometric error for corresponding epipolar tangencies. | 43 |
| 3.5 | Estimation of epipolar geometry from profiles: projective case. | 44 |
| 3.6 | Estimation of epipolar geometry from profiles: affine case. | 44 |
| 3.7 | Geometric representation of the trilinear constraint. | 50 |
| 4.1 | Symmetry transformations related to the images of a surface of revolution. | 64 |
| 4.2 | Symmetry of the eipoles with respect to the image of the axis of rotation. | 68 |
| 4.3 | Harmonic homology as a plane induced homography. | 71 |
| 4.4 | Harmonic homology estimated from bitangents. | 73 |
| 4.5 | Transfer of epipolar lines via harmonic homology. | 74 |

| | | |
|------|--|-----|
| 4.6 | Geometric error as a function of the orientation of the epipolar lines. | 75 |
| 4.7 | Geometric error of corresponding epipolar lines. | 77 |
| 4.8 | Real image sequences for estimation of epipolar geometry. | 79 |
| 4.9 | Surface of revolution generated by rotating object. | 80 |
| 4.10 | Robust fitting of the horizon to a cloud of epipoles. | 81 |
| 4.11 | Camera configurations for the real image sequences | 81 |
| 4.12 | Estimated angles of rotation between successive cameras. | 82 |
| 4.13 | Triangulation of profiles. | 84 |
| 4.14 | Sampling of points along profile according to the epipolar constraint. | 85 |
| 4.15 | Reconstruction of a vase. | 86 |
| 4.16 | Reconstruction of a head model. | 87 |
| 5.1 | Stereo rig with intersecting optical axes. | 100 |
| 5.2 | Measures of criticality of the motion. | 102 |
| 5.3 | Cameras and points used in the self-calibration experiment with synthetic data. | 107 |
| 5.4 | RMS error in the estimation of focal lengths 600. | 108 |
| 5.5 | RMS error in the estimation of focal lengths 1067. | 108 |
| 5.6 | RMS error in the estimation of focal lengths 1533. | 109 |
| 5.7 | RMS error in the estimation of focal lengths 2000. | 110 |
| 5.8 | Tracking of variable focal length. | 110 |
| 5.9 | Images used on experiment with real data. | 111 |
| 5.10 | Estimated values of focal length from different algorithms. | 112 |
| 5.11 | Normalised error on focal lengths. | 113 |
| 5.12 | Estimation of principal point. | 113 |
| 5.13 | Images for outdoor reconstruction. | 115 |

5.14 3D reconstruction of outdoor scene. 116

Glossary

[...] In geometry (which is the only science that it hath pleased God hitherto to bestow on mankind), men begin at settling the significations of their words; which significations [...] they call definitions, and place them in the beginning of their reckoning.

Thomas Hobbes (1588–1679), *Leviathan*, part I, chapter 4, 1651.

Sets

| | |
|-----------------|---|
| $A \times B$ | Cartesian product of the sets A and B . |
| $A \setminus B$ | Complement of B with respect to A . |
| C^m | Set of m -differentiable manifolds. |
| \mathbb{R} | Set of real numbers. |
| \mathbb{R}^n | n -dimensional Euclidean space. |
| \mathcal{P}^n | n -dimensional projective space. |

Points, vectors and matrices

| | |
|----------------|---|
| \mathbf{x}_c | Cartesian vector in 2-dimensional space. |
| \mathbf{X}_c | Cartesian vector in n -dimensional space. |
| \mathbf{x} | Homogeneous vector in 2-dimensional space. |
| \mathbf{X} | Homogeneous vector in n -dimensional space. |
| \mathbb{I} | Identity matrix. |
| \mathbb{O} | Null matrix. |
| x | Point in 2-dimensional space. |
| X | Point in n -dimensional space. |

Operators

| | |
|----------------------------|---|
| \mathbf{A}^* | Adjoint of matrix \mathbf{A} . |
| $\det(\mathbf{A})$ | Determinant of matrix \mathbf{A} . |
| \mathbf{A}^{-1} | Inverse of matrix \mathbf{A} . |
| \otimes | Kronecker product. |
| $\mathcal{LN}(\mathbf{A})$ | Left null space of the matrix \mathbf{A} . |
| \mathbf{A}^\dagger | Moore-Penrose inverse of matrix \mathbf{A} . |
| $\rho(\mathbf{A})$ | Rank of matrix \mathbf{A} . |
| $\mathcal{RN}(\mathbf{A})$ | Right null space of the matrix \mathbf{A} . |
| $\text{svd}(\mathbf{A})$ | Singular value decomposition of matrix \mathbf{A} . |
| \mathbf{A}^T | Transpose of matrix \mathbf{A} . |
| $\text{vec}(\mathbf{A})$ | vec operator. |
| \wedge | Wedge product. |

Variables

| | |
|--------------|------------------------------|
| \mathbf{E} | Essential matrix. |
| \mathbf{F} | Fundamental matrix. |
| \mathbf{K} | Intrinsic parameters matrix. |
| \mathbf{P} | Projective camera matrix. |
| \mathbf{R} | Rotation matrix. |
| \mathbf{t} | Translation vector. |
| T_i^{jk} | Trifocal tensor. |

Chapter 1

Introduction

1.1 Motivation

Two problems were tackled in this thesis: estimation of epipolar geometry from profiles (section 3.3 and chapter 4) and camera self-calibration (chapter 5). They both have a long history in computer vision, and much has been written about them (an extensive bibliography to the subjects can be found in the corresponding chapters). The amount of attention drawn by these topics is justified, for they are central to the solution of a key problem in computer vision, which is *the recovery of camera motion and 3D scene structure from video sequences*.

1.1.1 Shape from X

Several approaches have been used to attack these problems. In *shape from shading* the 3D structure of an object can be inferred from a single image by using physical models that link the intensity of the light reflected from a surface to properties of the surface material and the orientation of the surface with respect to ex-

ternal light sources [74, 75, 160]. *Shape from texture* algorithms intend to recover both 3D structure and motion from the relative deformation of homogeneous texture elements on a surface, denoted *texels* [157, 86, 105]. Another possibility is *shape from specularity*, which uses specular reflection as a clue for scene structure [88, 11, 119]. In *shape from focus* the blurring of the image at different distances from the camera offers information about surface depth; a common variation of this method is obtained when the focus of the camera is deliberately changed to blur the image, a technique known as *shape from defocus* [120, 144, 117, 22].

1.1.2 Geometry-Based Algorithms

Although extremely useful for specific applications, none of the methods mentioned above is as flexible as or produces results with accuracy comparable to stereo or multicamera geometry-based systems [41, 70]. A common feature of this class of techniques is the computation of projective invariants related to the set of cameras in the form of the *epipolar geometry*, *trilinear constraints*, or generic *multiview camera relations* (chapters 3 and 4), and the use of intrinsic camera parameters, which can be either precomputed or computed after the estimation of the epipolar geometry (chapter 5).

Direct versus Feature Based Methods. In order to obtain these projective invariants it is necessary to match elements that can be seen simultaneously in images. If the matched elements are whole areas of the image, or any other image quantity directly measurable at each pixel, the technique is dubbed a *direct method* [77]. If, alternatively, a sparse set of matched features is used, such as points, lines or contours, the method is denoted to be *feature based* [149]. For a comparison between

the two approaches see [3].

1.1.3 Structure and Motion from Profiles

To estimate the structure and the motion of smooth textureless surfaces, neither direct nor feature based methods using points or lines are appropriate. In this situation the *profile* of the surface is the most dominant feature in the image, and it offers important information for determining both the shape and the motion of the surface [28]. Moreover, even when the matching of other features is possible, the information provided by the profiles should not be neglected. One of the objectives of this work was, therefore, the development of practical algorithms for estimating the epipolar geometry of a sequence of images from profiles.

1.1.4 Camera Self-Calibration

The invariants encoded by the epipolar geometry or other multiview camera relations are projective ones, i.e., they allow the recovery of scene structure and camera motion only up to a projective transformation [40, 63]. To upgrade this projective transformation into an Euclidean one, or at least into a similarity transformation, it is necessary to obtain the intrinsic parameters of the cameras that acquired the images. If the Euclidean structure of the scene is already known, this knowledge can be used to compute the camera parameters, in a technique called *camera calibration* [49, 155]. For unknown scenes, a more general approach, denoted *camera self-calibration* [109], must be employed.

Critical Motions for Self-Calibration. There is a rich pool of algorithms designed to solve the self-calibration problem. However, there are theoretical limita-

tions to what these algorithms can achieve, due to the occurrence of *critical-motions* (section 5.4.2), which are camera configurations that involve the self-calibration up to the point of rendering it impossible. Until now, self-calibration and the analysis of critical motions have been done independently, and this work attempted to close this gap by developing a self-calibration algorithm with built-in critical motion detection.

1.2 Approach

Although projective geometry was the main mathematical tool used to develop the ideas presented in this thesis, greater emphasis was given to a rigorous analysis than to pure geometric insight whenever that was possible. When needed, geometric arguments were only complementary to the analytical ones, for it is the opinion of the author that intuition is not a substitute for formalism, but only an aid and an inspiration to the development of a scientific work.

As an engineering discipline, computer vision adopts the common paradigm of representing the world through models that are complex enough to allow useful descriptions of real phenomena, but nevertheless simple enough to render the analysis of these phenomena mathematically tractable. From this perspective, it is more beneficial to think of models as being useful or not, instead of simply qualifying them as right or wrong. Consider the affine camera model [1]. Strictly speaking, its use as a description of real cameras is, most of the time, wrong. However, it is well known that the advantages obtained in employing it can, in some circumstances, surpass possible gains attained from adopting the more complex projective camera model [70]. The same is true for ignoring lens distortion or other optical aberrations.

1.2.1 Models and Approximations Adopted

The projective or pinhole camera, which is probably the most important model used in geometric computer vision, was adopted almost everywhere in this work. An exception can be found in section 3.3, which made use of affine approximation to solve the problem of estimating the epipolar geometry from profiles under general camera motion. Again in section 3.3 and also in chapter 4, smooth objects were modelled as C^1 surfaces, i.e., subsets \mathcal{S} of \mathbb{R}^3 such that for each point $\mathbf{X} \in \mathcal{S}$ there exists a *proper patch* in \mathcal{S} whose image contains a neighborhood of \mathbf{X} in \mathcal{S} [118]. Still in chapter 4, the overlapping of a finite number of images of a rotating object was approximated to the image of a surface of revolution.

Validity of the Models. Most of the success of geometric computer vision depends on the validity of the pinhole camera model. Although some applications demand the correction of lens radial distortion [32], for the range of focal lengths used here the pinhole camera was an entirely appropriate model. The conditions for the validity of the affine camera model are well-known: narrow field of view and large focal length [73].

Since a B-spline was fitted to the profiles of the objects used in this work, it was required to model these objects as C^1 surfaces or at least piecewise C^1 surfaces (if two control points are allowed to coincide). This condition is violated by surfaces with fractal patches, which, although continuous everywhere, are not differentiable anywhere in the fractal patch. This situation roughly corresponds to highly spiky surfaces, for which, therefore, the algorithms presented here should not be employed.

Finally, for the purposes of this work the assumption that the overlapping of

a finite number of images of a rotating object can be regarded as the image of a surface of revolution was empirically verified as valid for angles of rotation between successive snapshots of up to 15° .

1.3 Contributions

The main contributions of this thesis are:

- a solution to the problem of motion estimation from profiles in the case of circular motion, based on symmetry properties of images of surfaces of revolution; this is the first solution to the problem of motion estimation from profiles that is both practical to implement and general enough to allow the Euclidean reconstruction of the scene;
- a novel self-calibration technique, based on the Huang and Faugeras constraints for essential matrices; the method introduced is flexible, for it can be employed for image sequences with any combination of fixed and varying intrinsic parameters, and it also takes into account the proximity of the relative motion of camera pairs to a critical configuration for self-calibration, greatly improving its robustness.

Minor contributions include:

- a specialisation of a generic method for estimating epipolar geometry from profiles to the affine case, aided by a new parameterisation of the affine fundamental matrix that is both minimal and general;
- a novel minimal parameterisation of the trifocal tensor that can be easily extended to provide minimal parameterisations for multiview tensors of any de-

gree.

1.4 Outline of the Thesis

Chapter 2. This chapter presents a brief review of some topics of projective geometry used elsewhere in this thesis. It introduces homogeneous coordinates and describes the representation of geometric primitives such as points, lines, planes and surfaces in such coordinates. It then analyses the imaging process carried out by a projective camera in the context of general projective transformations. Finally, it discusses some issues related to the projection of planes and surfaces.

Chapter 3. An overview of epipolar geometry and multiview camera relations is presented in chapter 3. It starts with the derivation of the fundamental matrix, discussing its degrees of freedom, a plane plus parallax representation and the essential matrix. It also mentions methods for computing the fundamental matrix from point matches and succinctly describes aspects of the epipolar geometry of affine cameras. The summary of epipolar geometry is closed with a quick presentation of epipolar geometry and smooth surfaces, and describes an algorithm for estimating affine epipolar geometry from profiles, presenting results of implementation on real data.

In addition, this chapter brings a summary of geometry of multiple cameras with emphasis on the trifocal tensor. It derives a novel minimal parameterisation of the trifocal tensor and discusses its computation. Furthermore, the chapter specialises the discussion of the trifocal tensor to the affine case, and, at last, mentions some aspects of multilinear camera relations.

Chapter 4. This chapter describes in detail one of the main contributions of this work, which is the development of an algorithm for estimating the epipolar geometry of a turntable sequence based solely on profiles. It brings an extensive bibliographic review on motion estimation and reconstruction from profiles, followed by a discussion on symmetry of images of surfaces of revolution. A novel parameterisation of the fundamental matrix is derived, which is then used to develop an algorithm to estimate the epipolar geometry of the cameras. The implementation of this algorithm for real images is also shown in chapter 4, and examples of 3D reconstruction using the motion obtained from the estimation of the epipolar geometry are presented. The results obtained in this chapter are the product of close collaboration carried out with Kwan-Yee Kenneth Wong.

Chapter 5. This chapter begins with a summary of current and past literature on self-calibration. It then proceeds to review some properties of the essential matrix, the Huang and Faugeras constraints, which are used to develop a linear algorithm for computation of focal lengths given the fundamental matrices related to a set of cameras. An analysis of this algorithm provides a method for detecting critical camera motions, and the solution provided by the linear algorithm is then refined through a nonlinear procedure that also takes into account critical motion configurations. Experimental results for both synthetic and real data are shown.

Chapter 6. The conclusion presents a summary of the work and points to directions for future research on the topics discussed in this thesis.

Chapter 2

Projective Geometry

Let no one enter who does not know geometry.

Plato (429–347 B.C.),
Elias Philosophus, *Aristotelis Categoriae Comentariorum*.

2.1 Introduction

Projective geometry is the invariant theory of the group of projective transformations [133, p. 41]. If Euclidean geometry is interpreted as the geometry of the straight edge and compass, projective geometry is the geometry of the straight edge alone [31].

This chapter presents a brief review of some fundamentals of projective geometry that will be necessary throughout the remainder of this dissertation. No attempt has been made to provide a comprehensive survey, which can be found in many good works on projective geometry [133, 140, 30, 31] and computer vision [8, 115, 41, 70]. Although the treatment of the subject presented here is non-standard and some rather advanced topics are covered, a reader proficient in projective geometry may want to skip this chapter and proceed to chapter 3.

2.2 Homogeneous Coordinates

Homogeneous coordinates are an analytical tool most suitable for tackling problems in projective geometry, playing a role equivalent to the one Cartesian coordinates play to Euclidean geometry [108]. Consider a point X in the n -dimensional space with Cartesian coordinates given by the n -tuple $(X_1, X_2, \dots, X_n) \in \mathbb{R}^n$. The expression of X in homogeneous coordinates is the set of $(n + 1)$ -tuples $\{w(X_1, X_2, \dots, X_n, 1) \mid w \in \mathbb{R} \setminus \{0\}\}$. Conversely, given the homogeneous coordinates $\{w(X_1, X_2, \dots, X_n, X_{n+1}) \in \mathbb{R}^{n+1} \setminus \{(0, 0, \dots, 0, 0)\}, \forall w \neq 0\}$ of a point X in the n -dimensional space, the Cartesian coordinates of X will be given by $(X_1, X_2, \dots, X_n)/X_{n+1}$, if $X_{n+1} \neq 0$. If $X_{n+1} = 0$, the point X is said to be at infinity in direction (X_1, X_2, \dots, X_n) , and it cannot be represented in Cartesian coordinates.

A vectorial representation of X in the canonical basis and in homogeneous coordinates is given by the set of column vectors \mathbf{X} where

$$\mathbf{X} = \left\{ w [X_1 \dots X_n X_{n+1}]^T \mid w \in \mathbb{R} \setminus \{0\} \right\}. \quad (2.1)$$

The corresponding representation of X in Cartesian coordinates is given by \mathbf{X}_c where

$$\mathbf{X}_c = [X_1/X_{n+1} \dots X_n/X_{n+1}]^T, \quad (2.2)$$

if $X_{n+1} \neq 0$. Henceforth, there will be made no distinction between a point and its vectorial representation in the canonical basis, whether in homogeneous coordinates or not, except where explicitly mentioned.

From the above discussion it can be seen that the representation of a point in

homogeneous coordinates is actually a set, and the equation

$$\mathbf{X} = \mathbf{X}' \quad (2.3)$$

expresses, in fact, the equality between the sets $\{w[X_1 \ X_2 \ \dots \ X_n \ X_{n+1}]^T \ \forall w \in \mathbb{R} \setminus \{0\}\}$ and $\{w'[X'_1 \ X'_2 \ \dots \ X'_n \ X'_{n+1}]^T \ \forall w' \in \mathbb{R} \setminus \{0\}\}$. Interpreting equality between representations in homogeneous coordinates as an equality between sets is a formal way of avoiding having explicitly to write the free scale factors of the terms involved, or having to use the symbol \approx to indicate “equality up to a nonzero scale factor”, as it is commonly done [66]. Hereafter, the vectorial representation $\{w[X_1 \ X_2 \ \dots \ X_n \ X_{n+1}]^T \ \forall w \in \mathbb{R} \setminus \{0\}\}$ of an n -dimensional object X in homogeneous coordinates will be shortened to $[X_1 \ X_2 \ \dots \ X_n \ X_{n+1}]^T$.

Wedge Product and Wedge Operator. Let \mathbf{X}^i , $i = \text{I, II, } \dots, n$, be a set of n n -dimensional homogeneous vectors, and let \mathbb{X} be the $(n+1) \times n$ matrix given by $\mathbb{X} = [\mathbf{X}^{\text{I}} \ \mathbf{X}^{\text{II}} \ \dots \ \mathbf{X}^n]$. Finally, let \mathbb{X}_j be the matrix obtained from \mathbb{X} by removing its j -th row. The *wedge product* of \mathbf{X}^i , $i = \text{I, II, } \dots, n$, denoted $\mathbf{X}^{\text{I}} \wedge \mathbf{X}^{\text{II}} \wedge \dots \wedge \mathbf{X}^n$, is defined as

$$\mathbf{X}^{\text{I}} \wedge \mathbf{X}^{\text{II}} \wedge \dots \wedge \mathbf{X}^n = \left[\det(\mathbb{X}_1) \quad -\det(\mathbb{X}_2) \quad \dots \quad (-1)^{n+1} \det(\mathbb{X}_{n+1}) \right]^T. \quad (2.4)$$

Moreover, given the set \mathbf{X}^i , $i = \text{I, II, } \dots, n-1$, of $n-1$ n -dimensional homogeneous vectors, the *wedge operator* of \mathbf{X}^i , $i = \text{I, II, } \dots, n-1$, is defined as the multilinear (i.e., linear on each \mathbf{X}^i) mapping that takes \mathbf{X}^i , $i = \text{I, II, } \dots, n-1$ to the unique

$(n + 1) \times (n + 1)$ antisymmetric matrix $[\mathbf{X}^I \wedge \mathbf{X}^{II} \wedge \dots \wedge \mathbf{X}^{n-1}]_{\wedge}$ such that

$$[\mathbf{X}^I \wedge \mathbf{X}^{II} \wedge \dots \wedge \mathbf{X}^{n-1}]_{\wedge} \mathbf{X} = \mathbf{X}^I \wedge \mathbf{X}^{II} \wedge \dots \wedge \mathbf{X} \quad (2.5)$$

for all $\mathbf{X} \in \mathcal{P}^n$.

2.2.1 Lines and Planes in Homogeneous Coordinates

Consider a point x in 2D space with vectorial representation in Cartesian coordinates given by $\mathbf{x}_c = [x_1 \ x_2]^T$ and lying on a line l . Therefore \mathbf{x}_c satisfies an equation of the form

$$l_1 x_1 + l_2 x_2 + l_3 = 0, \quad (2.6)$$

where l_1, l_2 and l_3 are parameters defining the line. If x is represented in homogeneous coordinates, (2.6) can be rewritten as

$$\mathbf{l}^T \mathbf{x} = 0, \quad (2.7)$$

where $\mathbf{l} = [l_1 \ l_2 \ l_3]^T$ is the vectorial representation of l in homogeneous coordinates. In general, the equation of an n -dimensional plane π in homogeneous coordinates is given by

$$\boldsymbol{\pi}^T \mathbf{X} = 0, \quad (2.8)$$

where $\boldsymbol{\pi} = [\pi_1 \ \pi_2 \ \dots \ \pi_{n+1}]^T$.

The intersection \mathbf{x} of the lines \mathbf{l} and \mathbf{l}' can be conveniently expressed in homogeneous coordinates as $\mathbf{x} = \mathbf{l} \wedge \mathbf{l}'$. In general, the intersection \mathbf{X} of n n -dimensional

planes π^i , $i = \text{I, II, } \dots, n$ is given by the wedge product $\mathbf{X} = \pi^{\text{I}} \wedge \pi^{\text{II}} \wedge \dots \wedge \pi^n$.

The Duality Principle. From (2.7) and (2.8) it can be seen that, algebraically, the role of a line or more generally, of a hyperplane, is dual to that of a point when these objects are represented in homogeneous coordinates. Thus, a theorem T regarding hyperplanes and points has a dual theorem T' where the word “point” is substituted by “plane” and vice-versa, appropriate linguistic adjustments notwithstanding, as stated by the *duality principle* [133, p. 79]. An immediate consequence of this principle is that the n -dimensional plane π defined by the n points \mathbf{X}^i , $i = \text{I, II, } \dots, n$ is given by $\pi = \mathbf{X}^{\text{I}} \wedge \mathbf{X}^{\text{II}} \wedge \dots \wedge \mathbf{X}^n$.

2.2.2 Conics and Quadrics in Homogeneous Coordinates

Consider the equation

$$ax_1^2 + 2bx_1x_2 + 2cx_1 + dx_2^2 + 2ex_2 + f = 0 \quad (2.9)$$

of a conic C . In homogeneous coordinates (2.9) becomes

$$\mathbf{x}^T \mathbf{C} \mathbf{x} = 0, \quad (2.10)$$

where

$$\mathbf{C} = \begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix}. \quad (2.11)$$

The matrix \mathbf{C} is the representation of the conic C in homogeneous coordinates. In general, the equation of a n -dimensional quadric Q is given by

$$\mathbf{X}^T \mathbf{Q} \mathbf{X} = 0, \quad (2.12)$$

where \mathbf{Q} is a $(n + 1) \times (n + 1)$ symmetric matrix.

Given the quadric defined by the nonsingular matrix \mathbf{Q} , the set of planes π that satisfy the equation $\pi^T \mathbf{Q}^{-1} \pi = 0$ defines the *envelope* of the quadric \mathbf{Q} . Given a point \mathbf{X} and a quadric \mathbf{Q} , the plane $\pi = \mathbf{Q}\mathbf{X}$ is defined as the *polar* of \mathbf{Q} with respect to \mathbf{X} (figure 2.1).

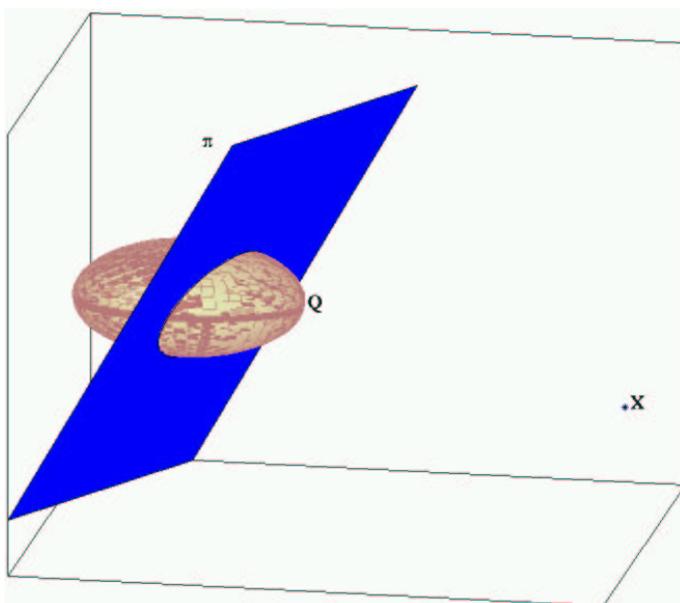


Figure 2.1. The plane π is polar to the quadric \mathbf{Q} with respect to the point \mathbf{X} .

Degenerate Cases If the matrix \mathbf{Q} of a quadric is singular, the quadric is said to be *degenerate*. Let $\rho(\mathbf{Q})$ be the rank of the quadric \mathbf{Q} . Particular cases of interest occur when:

- (i) $\rho(\mathbf{Q}) = 1$ — in this case the quadric corresponds to a hyperplane π , and it can be expressed as $\mathbf{Q} = \pi\pi^T$;
- (ii) $\rho(\mathbf{Q}) = 2$ — in this case the quadric corresponds to a pair of hyperplanes π and π' , and it can be expressed as $\mathbf{Q} = \pi\pi'^T + \pi'\pi^T$;

Given the $n - 1$ n -dimensional points \mathbf{X}^i , $i = \text{I, II, } \dots, n - 1$, the pair of planes tangent to the nondegenerate quadric \mathbf{Q} and containing the points \mathbf{X}^i , denoted as \mathfrak{P} , is given by

$$\mathfrak{P} = [\mathbf{X}^{\text{I}} \wedge \mathbf{X}^{\text{II}} \wedge \dots \wedge \mathbf{X}^{n-1}]_{\wedge}^T \mathbf{Q}^{-1} [\mathbf{X}^{\text{I}} \wedge \mathbf{X}^{\text{II}} \wedge \dots \wedge \mathbf{X}^{n-1}]_{\wedge}. \quad (2.13)$$

2.3 Projective Transformations

A linear operator applied over the homogeneous representation of a point defines a *projective transformation*. More formally, a projective transformation M from \mathcal{P}^n to \mathcal{P}^m is defined as

$$\begin{aligned} M : \mathcal{P}^n &\longrightarrow \mathcal{P}^m \\ \mathbf{X} &\longmapsto \mathbf{Y} = \mathbf{M}\mathbf{X}, \end{aligned} \quad (2.14)$$

where \mathbf{M} is a nonzero $(m + 1) \times (n + 1)$ matrix. Since \mathbf{X} and \mathbf{Y} are representations of points in homogeneous coordinates, the matrix \mathbf{M} in the definition of the projective transformation M can be considered as the equivalence class $\{s\mathbf{M}, \forall s \neq 0\}$. Therefore, the equality between the matrices \mathbf{M} and \mathbf{M}' associated with projective transformations M and M' corresponds to an equality between two sets. Hence-

forth, there will be no distinction made between the projective transformation M and the set $s\mathbf{M} \forall s \neq 0$ (shortened to \mathbf{M}) that defines the transformation.

Of special interest are the projective transformations from \mathcal{P}^3 to \mathcal{P}^3 , from \mathcal{P}^2 to \mathcal{P}^2 and from \mathcal{P}^3 to \mathcal{P}^2 . The bijective projective mappings $\mathbf{H}_{3D} : \mathcal{P}^3 \rightarrow \mathcal{P}^3$ and $\mathbf{H}_{2D} : \mathcal{P}^2 \rightarrow \mathcal{P}^2$ are a *3D projective transformation* and a *2D projective transformation*, respectively. There is a hierarchy of such transformations obtained by successive specialisation of the projective transformation to *affine*, *similarity*, and *Euclidean* transformations [42]. This can be understood by noting that any nonsingular 4×4 or 3×3 homogeneous matrix $\mathbf{H} = \begin{bmatrix} \mathbf{H}_1 & \mathbf{h}_2 \\ \mathbf{h}_3^T & h_4 \end{bmatrix}$ such that

$$\mathbf{H}_1^\dagger \mathbf{H}_1 \mathbf{h}_3 = \mathbf{h}_3 \quad (2.15)$$

where \mathbf{A}^\dagger denotes the Moore-Penrose inverse of \mathbf{A} , can be decomposed as

$$\mathbf{H} = \underbrace{\begin{bmatrix} \mathbb{I} & \mathbf{0} \\ \mathbf{p}^T & 1 \end{bmatrix}}_{\mathbf{H}_{pp}: \text{purely projective transformation}} \underbrace{\begin{bmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{0} & 1 \end{bmatrix}}_{\mathbf{H}_a: \text{affine transformation}} \quad (2.16)$$

and, moreover, the matrix \mathbf{H}_a in (2.16) can also be decomposed as

$$\mathbf{H}_a = \underbrace{\begin{bmatrix} \overbrace{\mathbf{T}_u}^{\text{upper triangular matrix}} & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix}}_{\mathbf{H}_{pa}: \text{purely affine transformation}} \underbrace{\begin{bmatrix} w \overbrace{\mathbf{R}}^{\text{rotation matrix}} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix}}_{\mathbf{H}_s: \text{similarity transformation}}. \quad (2.17)$$

The similarity transformation \mathbf{H}_s corresponds to a scaling (by the factor w), followed by an Euclidean transformation (rotation (\mathbf{R}) plus translation (\mathbf{t})). The treat-

ment shown above for the stratification of projective transformations generalises the one introduced in [62].

2.3.1 Projective Cameras

The imaging process produced by a projective camera can be interpreted as a sequence of three projective transformations [115, 41]:

- (i) a $\mathcal{P}^3 \rightarrow \mathcal{P}^3$ mapping from the 3D world coordinate system to the camera coordinate system, represented by the matrix \mathbf{H}_i given by

$$\mathbf{H}_i = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix}, \quad (2.18)$$

where \mathbf{R} is a 3×3 rotation matrix and \mathbf{t} is a 3×1 translation vector.

- (ii) a $\mathcal{P}^3 \rightarrow \mathcal{P}^2$ mapping from the 3D camera coordinate system to a 2D image plane, represented by the matrix \mathbf{H}_{ii} such that

$$\mathbf{H}_{ii} = \begin{bmatrix} \mathbb{I} & \mathbf{0} \end{bmatrix}, \quad (2.19)$$

- (iii) a $\mathcal{P}^2 \rightarrow \mathcal{P}^2$ mapping corresponding to a change of coordinates on the image plane represented by the matrix \mathbf{K} given by [41]

$$\mathbf{K} = \begin{bmatrix} \alpha & s & u \\ 0 & k\alpha & v \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \alpha & \alpha \cot \nu & u \\ 0 & k'\alpha / \sin \nu & v \\ 0 & 0 & 1 \end{bmatrix}, \quad (2.20)$$

where α is the camera *focal length*¹, $k' = k \sin \nu$ is the *aspect ratio*, $\nu = \text{acot}(s/\alpha)$ is the *skew angle*, s is the *skew*, and (u, v) are the coordinates of the *principal point* (for a geometric interpretation of these parameters, see [41]).

The overall mapping corresponding to the imaging process is therefore given by the matrix \mathbf{P} where

$$\mathbf{P} = \mathbf{K} [\mathbf{R} \ \mathbf{t}]. \quad (2.21)$$

The matrix \mathbf{P} is denoted *projective camera matrix*, and the matrix \mathbf{K} corresponds to the *matrix of intrinsic parameters*. The matrices \mathbf{R} and \mathbf{t} are jointly named matrices of *extrinsic or external parameters*. If the matrix \mathbf{K} is known, the camera is said to be *calibrated*. Hereafter, the expressions “the camera \mathbf{P} ” and “the intrinsic parameters \mathbf{K} ” should be read as “the camera with projective camera matrix given by \mathbf{P} ” and “the intrinsic parameters represented by the matrix \mathbf{K} ”, respectively. Observe that the camera \mathbf{P} given by (2.21) is not in the most general form of a full row rank 3×4 matrix, because the matrices \mathbf{K} and \mathbf{R} are always invertible. If this constraint is ignored, the general form of a projective camera matrix will be given by

$$\mathbf{P} = [\mathbf{M} \ \mathbf{e}']. \quad (2.22)$$

The point \mathbf{x} that results from applying the transformation corresponding to \mathbf{P} to a point \mathbf{X} — see figure 2.2 — is denoted the *projection* or *image* of \mathbf{X} on \mathbf{P} .

¹Strictly speaking, the term α is the product of the magnification factor by the focal length; however, since these factors cannot be decoupled, it is common practice to refer to their product simply as focal length.

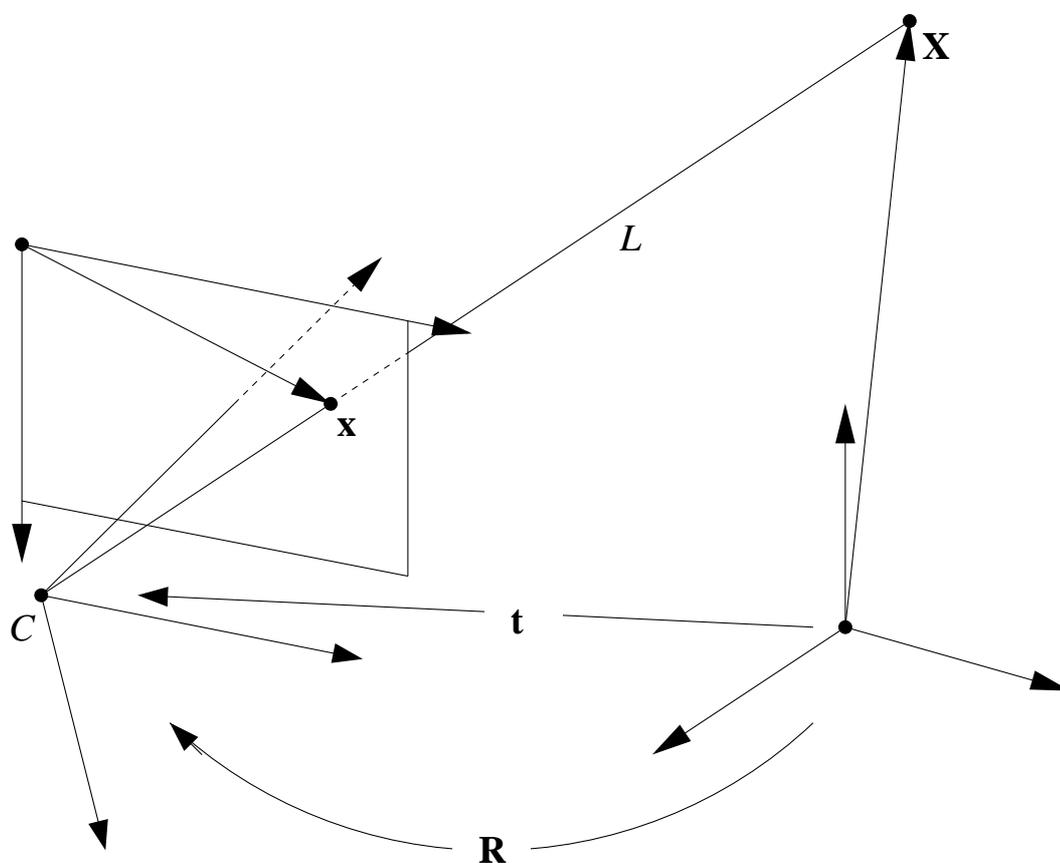


Figure 2.2. Sequence of projective transformations carried out by a projective camera: rotation R and translation t from world to camera coordinate system, projection from camera coordinate system to image plane, and transformation in image coordinates. The point C is the camera *optical centre*, and the line L is the *optical ray* associated with the point X .

The point C with homogeneous representation \mathbf{p}^\perp corresponding to the right null space of \mathbf{P} is the *optical centre* of the camera, and for each point \mathbf{x} in the camera coordinate system the set of 3D points $\mathbf{X}_L(w)$ given by $\mathbf{X}_L(w) = w\mathbf{p}^\perp + \mathbf{P}^\dagger\mathbf{x}$ defines a line L denoted as the *optical ray* or *ray* associated with the image point \mathbf{x} , as seen in figure 2.2.

2.3.2 Affine Cameras

Affine cameras are an important subclass of projective cameras. A camera \mathbf{P}_a is affine if [115]

$$\mathbf{P}_a = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ 0 & 0 & 0 & p_{34} \end{bmatrix} = \begin{bmatrix} \mathbf{p}_1^T & p_{14} \\ \mathbf{p}_2^T & p_{24} \\ \mathbf{0} & p_{34} \end{bmatrix}. \quad (2.23)$$

It has been shown [1, 115, 135] that if the depth of the scene in the direction of the optical axis is much smaller than the average distance of the scene to the camera, the projective camera can be approximated by an affine camera.

The right null space of \mathbf{P}_a is a vector of the form $[(\mathbf{p}_1 \wedge \mathbf{p}_2)^T \ 0]^T$. Therefore, the optical centre of an affine camera is at infinity in direction $\mathbf{p}_1 \wedge \mathbf{p}_2$. An affine camera can be obtained by substituting (2.19) by

$$\mathbf{H}_{ii_a} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2.24)$$

in the sequence of transformations involved in the imaging process of a projective camera, where \mathbf{H}_{ii_a} corresponds to the orthographic projection [115] of a point in 3D space onto the plane $z = 0$.

2.4 Projection of Planes and Surfaces

Let π be a plane in 3D space such that it contains the optical centre \mathbf{p}^\perp of a camera \mathbf{P} . Consider now a point \mathbf{X} on π , with image on \mathbf{P} given by \mathbf{x} . It follows that there

is a $w \in \mathbb{R}$ such that

$$\mathbf{X} = \mathbf{P}^\dagger \mathbf{x} + w \mathbf{p}^\perp, \quad (2.25)$$

and

$$\boldsymbol{\pi}^\top \mathbf{X} = 0, \quad (2.26)$$

$$\boldsymbol{\pi}^\top \mathbf{p}^\perp = 0. \quad (2.27)$$

Multiplying (2.25) on the left by $\boldsymbol{\pi}^\top$ and substituting (2.26) and (2.27) in the result, one obtains

$$\boldsymbol{\pi}^\top \mathbf{P}^\dagger \mathbf{x} = \mathbf{l}^\top \mathbf{x} = 0, \quad (2.28)$$

and therefore the image of the plane $\boldsymbol{\pi}$ is a line $\mathbf{l} = (\mathbf{P}^\top)^\dagger \boldsymbol{\pi}$. Analogously it is easy to show that the plane $\boldsymbol{\pi}$ whose image corresponds to a line \mathbf{l} is given by

$$\boldsymbol{\pi} = \mathbf{P}^\top \mathbf{l}. \quad (2.29)$$

2.4.1 Projection of Implicit Surfaces

Let \mathbf{Q} be a quadric in 3D space, where

$$\mathbf{Q} = \begin{bmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{b}^\top & c \end{bmatrix}, \quad (2.30)$$

and let \mathbf{P} be the camera $\mathbf{P} = [\mathbb{I} \ \mathbf{0}]$. Let now $\mathbf{X}_L(w)$ be the set of points on the optical ray L of an image point \mathbf{x} , i.e.,

$$\mathbf{X}_L(w) = [\mathbf{x}^T \ w]^T. \quad (2.31)$$

For each image point \mathbf{x} the 2nd degree equation in w

$$\mathbf{X}_L(w)^T \mathbf{Q} \mathbf{X}_L(w) = 0 \quad (2.32)$$

may have zero, one, or two real solutions. When (2.32) has only one solution, the line L will be tangent to \mathbf{Q} . Substituting (2.31) in (2.32) and imposing the constraint that the equation must have only one (real) solution, one obtains

$$\mathbf{x}^T (c\mathbf{A} - \mathbf{b}\mathbf{b}^T) \mathbf{x} = \mathbf{x}^T \mathbf{C} \mathbf{x} = 0, \quad (2.33)$$

which corresponds to the equation of a conic \mathbf{C} [34] (also [28, p. 70]). Observe that if the camera \mathbf{P} is not of the form $\mathbf{P} = [\mathbb{I} \ \mathbf{0}]$, but a generic 3×4 matrix with rank three and right null space \mathbf{p}^\perp , the conic can still be found by the method described if one considers the new pair camera-quadric $(\mathbf{P}', \mathbf{Q}')$ given by

$$([\mathbb{I} \ \mathbf{0}], \mathbf{H}^{-T} \mathbf{Q} \mathbf{H}^{-1}), \quad (2.34)$$

where $\mathbf{H} = [\mathbf{P}^\dagger \ \mathbf{p}^\perp]$, resulting in the same expression as found in [34]:

$$\mathbf{C}^* = \mathbf{P} \mathbf{Q}^* \mathbf{P}^T, \quad (2.35)$$

where \mathbf{A}^* denotes the adjoint of the matrix \mathbf{A} . The method used to derive (2.33) is in fact an application of elimination theory [103], which can be naturally used to extend the results to higher order surfaces [123, 124], as will be shown next.

Projecting Cubic and Quartic Surfaces. A d^{th} degree implicit surface in \mathcal{P}^n can be represented as a symmetric tensor \mathbf{T} of degree d , i.e.,

$$\sum_{i_1, i_2, \dots, i_d=1}^{n+1} T^{i_1, i_2, \dots, i_d} X_{i_1} X_{i_2} \dots X_{i_d} = 0. \quad (2.36)$$

For a camera $\mathbf{P} = [\mathbb{I} \ \mathbf{0}]$, consider the optical rays given by (2.31) associated with each image point \mathbf{x} . Substituting (2.31) in (2.36) one obtains a d^{th} degree polynomial $p(w)$ in w given by

$$p(w) = \sum_{j=0}^d \left[\binom{d}{j} \sum_{i_1, \dots, i_{d-j}=1}^n T^{i_1, \dots, i_{d-j}, n+1, \dots, n+1} x_{i_1} \dots x_{i_{d-j}} \right] w^j = 0. \quad (2.37)$$

The projection of \mathbf{T} will correspond to solutions of $p(w)$ with multiplicity greater than one. A necessary and sufficient condition for any polynomial $p(w)$ to have multiple roots is that the *resultant* of $p(w)$ and its derivative $p'(w)$ is zero [90]. The resultant of two polynomials is defined as the determinant of the *Sylvester matrix* of the polynomials [90, 23]. Therefore, an analytical expression for the implicit curve corresponding to the projection of \mathbf{T} can be derived by applying this condition to $p(w)$. It has been shown that if the degree of the implicit surface is d , the degree of the projection will be $d(d-1)$ [53]. Thus, the projection of implicit surfaces of degrees three and four will produce implicit curves of degrees six and twelve, respectively. As an example, figure 2.4 shows the twelfth-degree implicit curve corresponding to the projection of the fourth-degree implicit surface shown

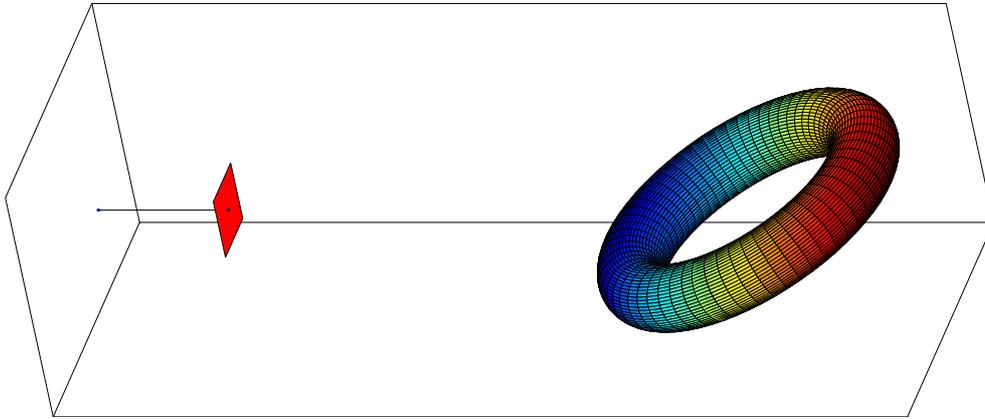


Figure 2.3. Torus corresponding to the fourth-degree implicit surface $x^4 + 2x^2y^2 + 2x^2z^2 - \frac{17}{2}x^2 + y^4 - \frac{17}{2}y^2 + z^4 + \frac{15}{2}z^2 + \left(\frac{225}{16}\right) = 0$ rotated by 68° about the X axis and translated 10 units along the Z axis.

in figure 2.3.

Implicit cubic surfaces are of particular interest for solid modelling. They are the surfaces of highest degree that, in general, have a rational parameterisation [132], and cubic surface patches can be easily joined to form more complex continuous forms [131].

2.5 Summary and Conclusions

This chapter presented a review of some aspects of projective geometry necessary for a better understanding of this dissertation. The main topics covered were:

Homogeneous coordinates — the presentation of homogeneous coordinates emphasised the interpretation of the homogeneous vector \mathbf{X} as a class of equivalence, simplifying the notation by avoiding the explicit representation of scale factors and making use of concepts from set theory; homogeneous representations for important geometric primitives, such as points, lines, planes and

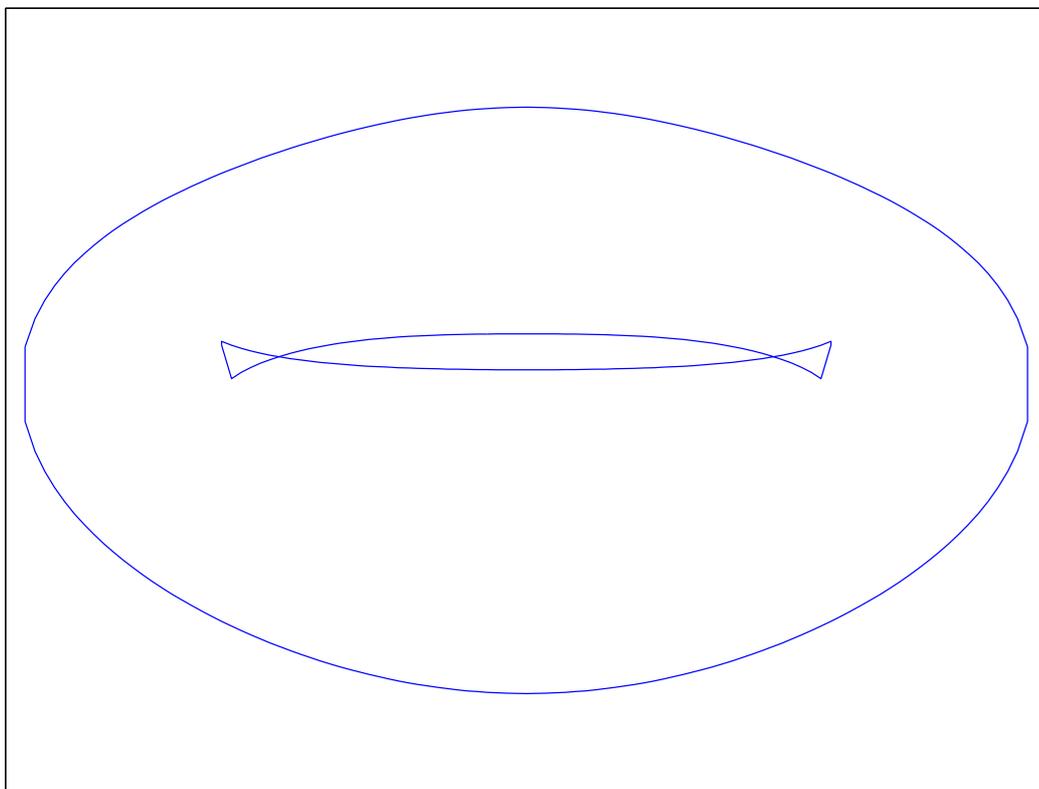


Figure 2.4. Projection of the torus shown on figure 2.3. The curve shown is an implicit twelfth-degree polynomial.

quadric surfaces were discussed;

Projective transformations — these were reviewed in a broad context, then specialised to the cases of particular interest to computer vision; the exposition on the hierarchy of transformations — projective, affine, Euclidean — generalises previous algebraic approaches; projective cameras were introduced in the same framework as general projective transformations;

Imaging of planes and implicit surfaces — finally, the imaging process of planes and implicit surfaces, regarded as a $\mathcal{P}^3 \rightarrow \mathcal{P}^2$ projective transformation, was briefly discussed and illustrated with an example of the projection of a 4th

order implicit surface.

Chapter 3

Epipolar Geometry and Multicamera Systems

This chapter presents an introduction to *epipolar geometry*, which is a set of geometric relations derived from stereo camera systems. Besides the general case, two particular configurations of the generic relations are also studied: one for affine cameras, and one for cameras viewing smooth surfaces. A new parameterisation of the *affine fundamental matrix* is introduced and applied to the problem of *estimation of affine epipolar geometry* from images of smooth surfaces. The chapter also briefly reviews some aspects of multicamera systems, in particular the *trifocal tensor* and its affine specialisation. A novel minimal parameterisation of the trifocal tensor is presented, which can be easily extended to parameterise multiview tensors of any degree, such as the *quadfocal tensor*.

3.1 Introduction

This section follows the approach of Zhang [164], which unifies the treatment of the fundamental matrix for projective and affine cameras.

3.1.1 The Fundamental Matrix

Consider a pair of cameras \mathbf{P} and \mathbf{P}' with distinct optical centres. The image of a point \mathbf{X} on each camera will be given by the points \mathbf{x} and \mathbf{x}' , where

$$\mathbf{x} = \mathbf{P}\mathbf{X}, \quad (3.1)$$

$$\mathbf{x}' = \mathbf{P}'\mathbf{X}. \quad (3.2)$$

Let \mathbf{p}^\perp be the homogeneous representation of the optical centre of \mathbf{P} . Therefore, the points \mathbf{X}_L on the line connecting \mathbf{p}^\perp to \mathbf{X} can be expressed in terms of the parameter w as

$$\mathbf{X}_L(w) = w\mathbf{P}^\dagger\mathbf{x} + \mathbf{p}^\perp, \quad (3.3)$$

For a particular choice w_0 of the parameter w , the point $\mathbf{X}_L(w_0)$ will be equal to \mathbf{X} . Therefore, $\mathbf{x}' = \mathbf{P}'\mathbf{X}_L(w_0)$, and

$$\mathbf{x}' = w_0\mathbf{P}'\mathbf{P}^\dagger\mathbf{x} + \mathbf{P}'\mathbf{p}^\perp. \quad (3.4)$$

Multiplying both sides of (3.4) by $\mathbf{x}'^T[\mathbf{P}'\mathbf{p}^\perp]_\wedge$, which is nonzero since the optical centres of \mathbf{P} and \mathbf{P}' are distinct, one obtains

$$\mathbf{x}'^T[\mathbf{P}'\mathbf{p}^\perp]_\wedge\mathbf{P}'\mathbf{P}^\dagger\mathbf{x} = 0. \quad (3.5)$$

This equation is the algebraic expression of the *epipolar constraint*, and the matrix

$$\mathbf{F} = [\mathbf{P}'\mathbf{p}^\perp]_\wedge\mathbf{P}'\mathbf{P}^\dagger \quad (3.6)$$

defines the *fundamental matrix* or *F matrix* of \mathbf{P} and \mathbf{P}' . From (3.5) and (3.6) it can be seen that the image \mathbf{x}' on \mathbf{P}' of a point with corresponding image \mathbf{x} on \mathbf{P} will satisfy $\mathbf{l}'^T\mathbf{x}' = 0$ where \mathbf{l}' is a line given by $\mathbf{l}' = \mathbf{F}\mathbf{x}$, denoted *epipolar line*. This configuration is symmetric, i.e., there is an epipolar line $\mathbf{l} = \mathbf{F}^T\mathbf{x}'$ on which the point \mathbf{x} lies. A geometric interpretation of the epipolar constraint can be seen in figure 3.1. The right and left null spaces of \mathbf{F} , denoted as $\mathcal{RN}(\mathbf{F})$ and $\mathcal{LN}(\mathbf{F})$, respectively, are given by

$$\mathcal{RN}(\mathbf{F}) = (\mathbf{P}'\mathbf{P}^\dagger)^{-1}\mathbf{P}'\mathbf{p}^\perp = \mathbf{e}, \quad (3.7)$$

$$\mathcal{LN}(\mathbf{F}) = \mathbf{P}'\mathbf{p}^\perp = \mathbf{e}'. \quad (3.8)$$

The homogeneous vectors \mathbf{e} and \mathbf{e}' are denoted right and left *epipoles*, and they represent the image of the optical centre of each camera as seen from the other (see figure 3.1). The points \mathbf{X} and the optical centres of \mathbf{P} and \mathbf{P}' define an *epipolar plane*.

Degrees of Freedom of F. A generic 3×3 matrix has nine degrees of freedom (dof), one for each of its entries. The fundamental matrix, however, is defined only

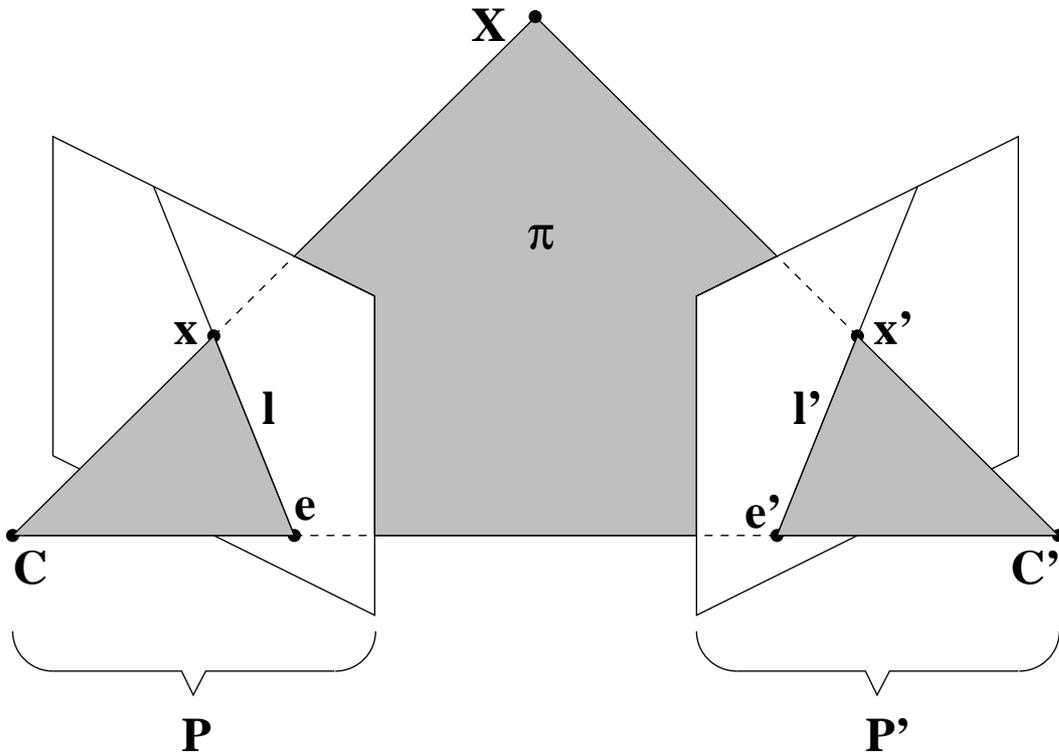


Figure 3.1. The *epipoles* e and e' correspond to the images of the optical centres C' and C , respectively. The point x' on camera P' corresponding to x on camera P must lie on the *epipolar line* l' . The points X , C and C' define the *epipolar plane* π .

up to a scale factor, since if a matrix \mathbf{F} satisfies the epipolar constraint for point correspondences between images on cameras P and P' , so does $w\mathbf{F}$, for any $w \neq 0$. Therefore, it can be imposed that $\|\mathbf{F}\|_N = 1$, where $\|\cdot\|_N$ indicates any matrix norm, e.g., Frobenius norm or p -norm [58]. Moreover, since \mathbf{F} corresponds to the product of a rank two matrix and a nonsingular matrix, as can be seen in (3.6), $\det(\mathbf{F}) = 0$. These two constraints reduce the number of dof of \mathbf{F} to seven.

3.1.2 Plane Plus Parallax Representation

It has been show [63] that the fundamental matrix relating any two cameras is invariant to the application of a common projective transformation to the world coordinate

system, i.e., the fundamental matrix $\mathbf{F}_{\mathbf{H}}$ associated with the pair of cameras \mathbf{P} and \mathbf{P}' is the same as the one associated with the cameras $\mathbf{P}_{\mathbf{H}} = \mathbf{P}\mathbf{H}$ and $\mathbf{P}'_{\mathbf{H}} = \mathbf{P}'\mathbf{H}$ for any full-rank matrix \mathbf{H} . This can be easily checked by substituting $\mathbf{P}_{\mathbf{H}}$ and $\mathbf{P}'_{\mathbf{H}}$ in (3.6)

$$\begin{aligned}\mathbf{F}_{\mathbf{H}} &= [\mathbf{P}'_{\mathbf{H}}\mathbf{p}_{\mathbf{H}}^{\perp}]_{\wedge}\mathbf{P}'_{\mathbf{H}}\mathbf{P}_{\mathbf{H}}^{\dagger} \\ &= [\mathbf{P}'\mathbf{H}\mathbf{H}^{-1}\mathbf{p}^{\perp}]_{\wedge}\mathbf{P}'\mathbf{H}(\mathbf{P}\mathbf{H})^{\dagger}\end{aligned}\quad (3.9)$$

and substituting the identity (where it is assumed that $\|(\mathbf{p}^{\perp})^{\dagger}\| = 1$)

$$\mathbb{I} = \mathbf{P}^{\dagger}\mathbf{P} + \mathbf{p}^{\perp}(\mathbf{p}^{\perp})^{\dagger}, \quad (3.10)$$

in (3.9), which then becomes

$$\begin{aligned}\mathbf{F}_{\mathbf{H}} &= [\mathbf{P}'\mathbf{p}^{\perp}]_{\wedge}\mathbf{P}'(\mathbf{P}^{\dagger}\mathbf{P} + \mathbf{p}^{\perp}(\mathbf{p}^{\perp})^{\dagger})\mathbf{H}(\mathbf{P}\mathbf{H})^{\dagger} \\ &= [\mathbf{P}'\mathbf{p}^{\perp}]_{\wedge}\mathbf{P}'\mathbf{P}^{\dagger}\mathbf{P}\mathbf{H}(\mathbf{P}\mathbf{H})^{\dagger} + [\mathbf{P}'\mathbf{p}^{\perp}]_{\wedge}\mathbf{P}'\mathbf{p}^{\perp}(\mathbf{p}^{\perp})^{\dagger}\mathbf{H}(\mathbf{P}\mathbf{H})^{\dagger} \\ &= [\mathbf{P}'\mathbf{p}^{\perp}]_{\wedge}\mathbf{P}'\mathbf{P}^{\dagger}\mathbb{I} + [\mathbf{P}'\mathbf{p}^{\perp}]_{\wedge}\mathbf{P}'\mathbf{p}^{\perp}\mathbf{0}^{\text{T}} \\ &= \mathbf{F}\end{aligned}\quad (3.11)$$

By choosing $\mathbf{H} = [\mathbf{P}^{\dagger} \ \mathbf{p}^{\perp}]$, the matrices $\mathbf{P}_{\mathbf{H}}$ and $\mathbf{P}'_{\mathbf{H}}$ become $\mathbf{P}_{\mathbf{H}} = [\mathbb{I} \ \mathbf{0}]$ and $\mathbf{P}'_{\mathbf{H}} = [\mathbf{P}'\mathbf{P}^{\dagger} \ \mathbf{P}'\mathbf{p}^{\perp}]$, or, denoting $\mathbf{P}'\mathbf{P}^{\dagger}$ by \mathbf{M} and using (3.8), $\mathbf{P}'_{\mathbf{H}} = [\mathbf{M} \ \mathbf{e}']$. In this notation the expression of the fundamental matrix is simplified to

$$\mathbf{F} = [\mathbf{e}']_{\wedge}\mathbf{M}. \quad (3.12)$$

Observe that if the matrix \mathbf{M} in (3.12) is substituted by any matrix

$$\mathbf{M}' = \mathbf{M} + \mathbf{e}'\mathbf{a}'^T, \quad (3.13)$$

the resulting fundamental matrix is the same [63]. However, since \mathbf{P} and \mathbf{P}' are full row rank, choosing $\mathbf{M} = \mathbf{P}'\mathbf{P}^\dagger$ ensures that \mathbf{M} is nonsingular.

Let now $\boldsymbol{\pi} = [\mathbf{n}^T \ -d]^T$ be the homogeneous representation of a given plane such that

$$d(d + \mathbf{n}^T\mathbf{M}^{-1}\mathbf{e}') \neq 0. \quad (3.14)$$

One can verify that this condition enforces the optical centres of $\mathbf{P}_{\mathbf{H}}$ and $\mathbf{P}'_{\mathbf{H}}$ not to lie on $\boldsymbol{\pi}$. Let now \mathbf{x} be the image of a point \mathbf{X} lying on $\boldsymbol{\pi}$. The optical ray associated with \mathbf{X} and the camera $\mathbf{P}_{\mathbf{H}}$ can be parameterised as $\mathbf{X}(w) = [\mathbf{x}^T \ w]^T$, where w is the free parameter. However, for a given $w = w_0$, $\mathbf{X}(w_0)$ must lie on $\boldsymbol{\pi}$, or $\boldsymbol{\pi}^T\mathbf{X}(w_0) = 0$, and therefore, $w_0 = \mathbf{n}^T\mathbf{x}/d$. Projecting $\mathbf{X}(w_0)$ on the camera $\mathbf{P}'_{\mathbf{H}}$ one obtains a point \mathbf{x}' given by

$$\begin{aligned} \mathbf{x}' &= \left(\mathbf{M} + \frac{\mathbf{e}'\mathbf{n}^T}{d} \right) \mathbf{x} \\ &= \mathbf{M}'\mathbf{x}, \end{aligned} \quad (3.15)$$

if \mathbf{a} in (3.13) is chosen as \mathbf{n}/d . The condition (3.14) prevents the matrix \mathbf{M}' from being singular, and therefore the images of points lying on $\boldsymbol{\pi}$ are related by the invertible map \mathbf{M}' , henceforth denoted *plane induced homography* [100, 102]. Note that if no reference is made to the fundamental matrix in the derivation of (3.15), the condition of distinct optical centres could be dropped, e.g., the cameras related

by a rotation about their common optical centre.

Still assuming distinct optical centres, the vector \mathbf{e}' can be seen to correspond to the vectorial representation of the Cartesian coordinates of the direction of the translation from $\mathbf{P}'_{\mathbf{H}}$ to $\mathbf{P}_{\mathbf{H}}$ in the coordinate system of $\mathbf{P}'_{\mathbf{H}}$ (the translation from $\mathbf{P}_{\mathbf{H}}$ to $\mathbf{P}'_{\mathbf{H}}$ in the coordinate system of $\mathbf{P}_{\mathbf{H}}$ is given by \mathbf{e}). Thus, the fundamental matrix is determined by an arbitrary plane induced homography and the relative translation between the cameras. This is the basis of the *plane plus parallax* approach [29, 92, 2, 78, 152] for analysing the geometry of multicamera systems.

3.1.3 The Essential Matrix

Assume now that the cameras \mathbf{P} and \mathbf{P}' have the form indicated in (2.21), i.e.,

$$\mathbf{P} = \mathbf{K} [\mathbf{R} \ \mathbf{t}] \quad (3.16)$$

$$\mathbf{P}' = \mathbf{K}' [\mathbf{R}' \ \mathbf{t}'] \quad (3.17)$$

Substituting (3.16) and (3.17) into (3.6), one obtains [102]

$$\mathbf{F} = \mathbf{K}'^{-\text{T}} [\mathbf{t}' - \mathbf{R}' \mathbf{R}^{\text{T}} \mathbf{t}]_{\wedge} \mathbf{R}' \mathbf{R}^{\text{T}} \mathbf{K}^{-1}. \quad (3.18)$$

Multiplying (3.18) on the left by \mathbf{K}'^{T} and on the right by \mathbf{K} results in a matrix \mathbf{E} , denoted *essential matrix* or *E matrix* [96], given by

$$\mathbf{E} = \mathbf{K}'^{\text{T}} \mathbf{F} \mathbf{K} \quad (3.19)$$

$$= [\mathbf{t}' - \mathbf{R}' \mathbf{R}^{\text{T}} \mathbf{t}]_{\wedge} \mathbf{R}' \mathbf{R}^{\text{T}} \quad (3.20)$$

Observe that the essential matrix depends only on the relative motion between the two cameras: $\mathbf{t}' - \mathbf{R}'\mathbf{R}^T\mathbf{t}$ and $\mathbf{R}'\mathbf{R}^T$ are the translation and rotation of \mathbf{P}' with respect to \mathbf{P} , respectively.

Degrees of Freedom of \mathbf{E} . Analogously to the fundamental matrix, the essential matrix is defined only up to a scale factor and its determinant is zero. However, for the decomposition given by (3.20) to be possible, two further constraints must be imposed [76], and, as a result, the matrix \mathbf{E} has only five dof. This fact can be exploited in self-calibration algorithms, as will be shown in chapter 5.

3.1.4 Computation of the Fundamental Matrix

Given two sets of images of the 3D points \mathbf{X}_i , $i = 1, 2, \dots, n$, obtained from the cameras \mathbf{P} and \mathbf{P}' , denoted as \mathbf{x}_i and \mathbf{x}'_i , respectively, each correspondence $(\mathbf{x}_i, \mathbf{x}'_i)$ provides, when substituted in (3.5), one equation on the entries of \mathbf{F} . If enough (at least seven) of such equations are available, a system of equations can be solved for \mathbf{F} . Different methods for solving these equations will produce different algorithms to compute the fundamental matrix.

Analytical Methods. Denoting $\mathbf{f} = \text{vec}(\mathbf{F})$ as the vector built out of the matrix \mathbf{F} by stacking its columns [104], the constraint $\mathbf{x}'_i{}^T \mathbf{F} \mathbf{x}_i = 0$ can be rewritten as

$$(\mathbf{x}_i^T \otimes \mathbf{x}'_i{}^T) \mathbf{f} = 0, \quad (3.21)$$

where \otimes indicates the Kronecker product. If seven correspondences are available, (3.21) can be augmented to a system of seven linear equations given by (with $N = 7$)

$$\begin{bmatrix} \mathbf{x}_1^T \otimes \mathbf{x}'_1{}^T \\ \mathbf{x}_2^T \otimes \mathbf{x}'_2{}^T \\ \vdots \\ \mathbf{x}_N^T \otimes \mathbf{x}'_N{}^T \end{bmatrix} \mathbf{f} = \mathfrak{X}\mathbf{f} = \mathbf{0}. \quad (3.22)$$

Let \mathbf{f}_1 and \mathbf{f}_2 be any two linearly independent vectors in the null space of \mathfrak{X} . Therefore, any solution of (3.22) can be expressed as $\mathbf{f} = w\mathbf{f}_1 + (1 - w)\mathbf{f}_2$. Substituting the general solution into the constraint $\det(\mathbf{F}) = 0$, one obtains a third degree polynomial on w that will have, in general, three solutions, each corresponding to a different fundamental matrix [63, 164].

For noise-free data the rank of \mathfrak{X} will be seven even for $N > 7$. However, the detection of image features is not a perfect process. There are a number of algorithms for feature detection, whether the features are points or corners [60, 59, 139], edges [107, 16, 17, 38], contours [83, 26, 36, 21, 12] etc. In the case of corners, the effect of noise is modelled by considering that the true position of the points is disturbed by an additive error with Gaussian distribution. Denoting the singular value decomposition (svd) of \mathfrak{X} as $\text{svd}(\mathfrak{X}) = \mathbf{U}_x \mathbf{D}_x \mathbf{V}_x^T$, with $\mathbf{V}_x = [\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_9]$, the least-squares solution of (3.22) becomes $\mathbf{f} = \mathbf{v}_9$. However, the fundamental matrix \mathbf{F} such that $\mathbf{f} = \text{vec}(\mathbf{F})$ does not necessarily fulfill the criteria $\det(\mathbf{F}) = 0$. To obtain a matrix \mathbf{F}' that satisfies this constraint, one can compute the singular value decomposition of \mathbf{F} , obtaining $\mathbf{F} = \mathbf{U}_F \mathbf{D}_F \mathbf{V}_F^T$, and make $\mathbf{F}' = \mathbf{U}_F \mathbf{D}'_F \mathbf{V}_F^T$, where \mathbf{D}'_F is obtained from \mathbf{D}_F by zeroing the last element in the diagonal of \mathbf{D}_F .

A common feature to analytical methods is that they minimise *algebraic error* [100, 68], which is an error measure that does not have a geometric or physical interpretation. Under the assumption of isotropic and homogeneous Gaussian noise, best solutions can be achieved by minimising the *geometric error* [100, 68, 164], which corresponds to the average of the Euclidean distance between each point \mathbf{x}' and the epipolar line $\mathbf{l}' = \mathbf{F}\mathbf{x}$, as shown in figure 3.2.

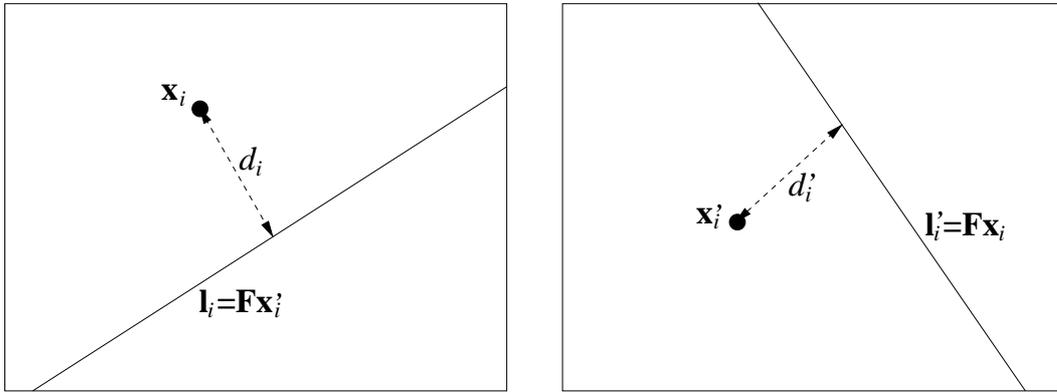


Figure 3.2. For each point correspondence $(\mathbf{x}, \mathbf{x}')$ the epipolar lines $\mathbf{l}' = \mathbf{F}\mathbf{x}$ and $\mathbf{l} = \mathbf{F}\mathbf{x}'$ can be computed. The square of the geometric error for the pair $(\mathbf{x}, \mathbf{x}')$ is then given by $\epsilon_i^2 = (d_i^2 + d_i'^2)/2$, and the average geometric error will be $\frac{1}{N} \sum_{i=1}^N \epsilon_i^2$.

Iterative Methods. In order to minimise the geometric error when estimating the epipolar geometry, it is necessary to make use of iterative procedures, since this formulation does not allow for an analytical solution to the problem. Given the pair of corresponding points $(\mathbf{x}_i, \mathbf{x}_i')$ and a fundamental matrix \mathbf{F} , let $\mathbf{l}' = \mathbf{F}\mathbf{x} = [l'_1 \ l'_2 \ l'_3]^T$ and $\mathbf{l} = \mathbf{F}\mathbf{x}' = [l_1 \ l_2 \ l_3]^T$ be the associated epipolar lines. Thus, the square of the geometric error of this configuration will be given by [102, 164]

$$\epsilon_i^2 = \left(\frac{1}{l_1^2 + l_2^2} + \frac{1}{l'_1{}^2 + l'_2{}^2} \right) \frac{(\mathbf{x}'\mathbf{F}\mathbf{x})^2}{2}. \quad (3.23)$$

If N correspondences are available, the average geometric error for the configuration will be

$$E = \frac{1}{N} \sum_{i=1}^N \epsilon_i^2. \quad (3.24)$$

The problem is then to find a 3×3 , rank two matrix \mathbf{F} that minimises the cost function (3.24). This can be achieved by using standard numerical methods [99, 126] such as Newton-Rhapson or Levenberg-Marquardt and an appropriate parameterisation of the fundamental matrix. The most natural parameterisation, where $(\mathbf{F})_{ij} = f_{ij}$, $i, j = 1, 2, 3$, suffers from two problems: first it is not minimal, i.e., it involves more parameters than the number of dof of \mathbf{F} ; second, it does not ensure the condition $\det(\mathbf{F}) = 0$. Two other possibilities were discussed in [102]:

$$\mathbf{F} = \begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_1 a_7 + a_4 a_8 & a_2 a_7 + a_5 a_8 & a_3 a_7 + a_6 a_8 \end{bmatrix} \quad (3.25)$$

and

$$\mathbf{F} = \begin{bmatrix} b & a & -ay - bx \\ -d & -c & cy + dx \\ dy' - bx' & cy' - ax' & -cyy' - dy'x + ayx' + bxx' \end{bmatrix}. \quad (3.26)$$

Both parameterisations are not minimal, since they involve eight parameters. Moreover, they are also not general, for they cannot represent configurations in which the epipoles are at infinity. As will be shown in section 3.2, this an important particular case of the epipolar geometry. An alternative parameterisation is introduced in

the appendix A, which, although not general, is minimal and does not preclude the epipoles from being at infinity. The gradient and the Hessian of E in (3.24) for this parameterisation are also given in the appendix A.

A thorough evaluation of several analytical and iterative methods for estimating the epipolar geometry from point matches can be found in [45, 102, 65, 164].

3.2 Affine Epipolar Geometry

For affine cameras, the epipolar geometry has some special properties. These properties can be derived by using different approaches, such as Grassman-Cayley algebra [46], direct substitution of the affine cameras onto the expression for the fundamental matrix [115], or projective geometry [110]. The latter method is more in the spirit of this work, and therefore it will be discussed here in more detail.

3.2.1 The Affine Fundamental Matrix

Consider two affine cameras $\mathbf{P}_{1,a}$ and $\mathbf{P}_{2,a}$. Without loss of generality, it can be imposed that

$$\mathbf{P}_{1,a} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (3.27)$$

If $\mathbf{P}_{1,a}$ has a different format, there is an affine transformation that when applied to both cameras will transform $\mathbf{P}_{1,a}$ to the form of (3.27). As discussed in subsection 3.1.2, the fundamental matrix \mathbf{F}_a related to $\mathbf{P}_{1,a}$ and $\mathbf{P}_{2,a}$ is invariant to the application of such a transformation. Consider now the projective transformation

\mathbf{H} given by

$$\mathbf{H}_q = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \quad (3.28)$$

The transformation \mathbf{H}_q is a very particular projective transformation. It does not satisfy condition (2.15), and it can be used to produce *quasi-affine reconstructions* [67]. By applying this transformation to both $\mathbf{P}_{1,a}$ and $\mathbf{P}_{2,a}$, the result is a new pair of matrices given by

$$\begin{aligned} \mathbf{P}_{1,a}\mathbf{H} &= [\mathbb{I} \quad \mathbf{0}], \\ \mathbf{P}_{2,a}\mathbf{H} &= \begin{bmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ 0 & 0 & \times & 0 \end{bmatrix}, \end{aligned} \quad (3.29)$$

where \times represents a possibly nonzero value. Substituting (3.29) in (3.12), one obtains

$$\mathbf{F}_a = \begin{bmatrix} 0 & 0 & \times \\ 0 & 0 & \times \\ \times & \times & 0 \end{bmatrix} \begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ 0 & 0 & \times \end{bmatrix} = \begin{bmatrix} 0 & 0 & \times \\ 0 & 0 & \times \\ \times & \times & \times \end{bmatrix}. \quad (3.30)$$

Therefore, a natural parameterisation for the *affine fundamental* matrix is

$$\mathbf{F}_a = \begin{bmatrix} 0 & 0 & a \\ 0 & 0 & b \\ c & d & e \end{bmatrix}, \quad (3.31)$$

where $(a, b, c, d, e) \in \mathbb{R}^5$. Since the overall scale factor is not important, the affine fundamental matrix has only four dof. Therefore, the parameterisation in (3.31) is not minimal, since it has five parameters. It could be argued that to turn (3.31) into a minimal parameterisation it would be enough to make any of the entries a , b , c , d or e equal to one. However, the resulting parameterisation would not be general, for it would prevent the fixed parameter from being zero, which is a perfectly valid situation. A better alternative is to express (3.30) as

$$\mathbf{F}_a = \begin{bmatrix} 0 & 0 & \sin \psi' \\ 0 & 0 & -\cos \psi' \\ \alpha \sin \psi & -\alpha \cos \psi & \beta \end{bmatrix}, \quad (3.32)$$

where $\alpha \in \mathbb{R} \setminus \{0\}$, $\beta \in \mathbb{R}$ and $(\psi, \psi') \in [-\pi, \pi) \times [-\pi, \pi)$. The right and left null spaces of \mathbf{F}_a are

$$\mathcal{RN}(\mathbf{F}_a) = [\cos \psi \ \sin \psi \ 0]^T; \quad (3.33)$$

$$\mathcal{LN}(\mathbf{F}_a) = [\cos \psi' \ \sin \psi' \ 0]^T. \quad (3.34)$$

Therefore, ψ and ψ' are the directions of the right and left epipoles, respectively. This parameterisation is similar to the one introduced in [111], and it can be shown that, if d is the distance of an epipolar line to the origin of the coordinate system on

the right image, the distance of the corresponding epipolar line to the origin of the coordinate system on the left image is $\alpha d - \beta$. Thus, the parameterisation shown in (3.32) is minimal, general, and it allows for a geometric interpretation of each one of the parameters ψ , ψ' , α and β .

3.3 Epipolar Geometry and Smooth Surfaces

A detailed exposition of estimation of epipolar geometry from images of smooth surfaces is given in the chapter 4. This section considers only some background material and results related to the affine epipolar geometry of smooth surfaces.

Consider a surface \mathcal{S} of type C^1 , as defined in section 1.2.1, viewed by two pinhole cameras \mathbf{P}_1 and \mathbf{P}_2 . The following definitions are presented as a quick review [28]:

- a *contour generator* associated with the surface \mathcal{S} and the camera \mathbf{P}_1 corresponds to the space curve $\mathcal{C} \subset \mathcal{S}$ such that for all points $\mathbf{C} \in \mathcal{C}$ the line passing through the optical centre of \mathbf{P}_1 and \mathbf{C} is tangent to \mathcal{S} at \mathbf{C} ;
- the image of the contour generator associated with the camera \mathbf{P}_1 is a *profile* or *apparent contour*;
- if two contour generators associated with the surface \mathcal{S} and the cameras \mathbf{P}_1 and \mathbf{P}_2 intersect, the points of intersection are denoted *frontier points*;
- the epipolar plane Π defined by the optical centers of the two cameras \mathbf{P}_1 and \mathbf{P}_2 and an associated frontier point \mathbf{X}_f is tangent to the surface \mathcal{S} at \mathbf{X}_f ;
- the epipolar lines corresponding to the epipolar plane Π are tangent to their associated profiles and are called *epipolar tangents*;

The tangent points of associated epipolar tangencies correspond to the images of the same point on the surface \mathcal{S} , namely, the frontier point. The above definitions are illustrated in figure 3.3. The key observation about the epipolar geometry of cameras viewing smooth surfaces is the following property [125, 28]:

Property 1 *Epipolar lines tangent to the profile in one image correspond to epipolar lines tangent to the profile at the other image.*

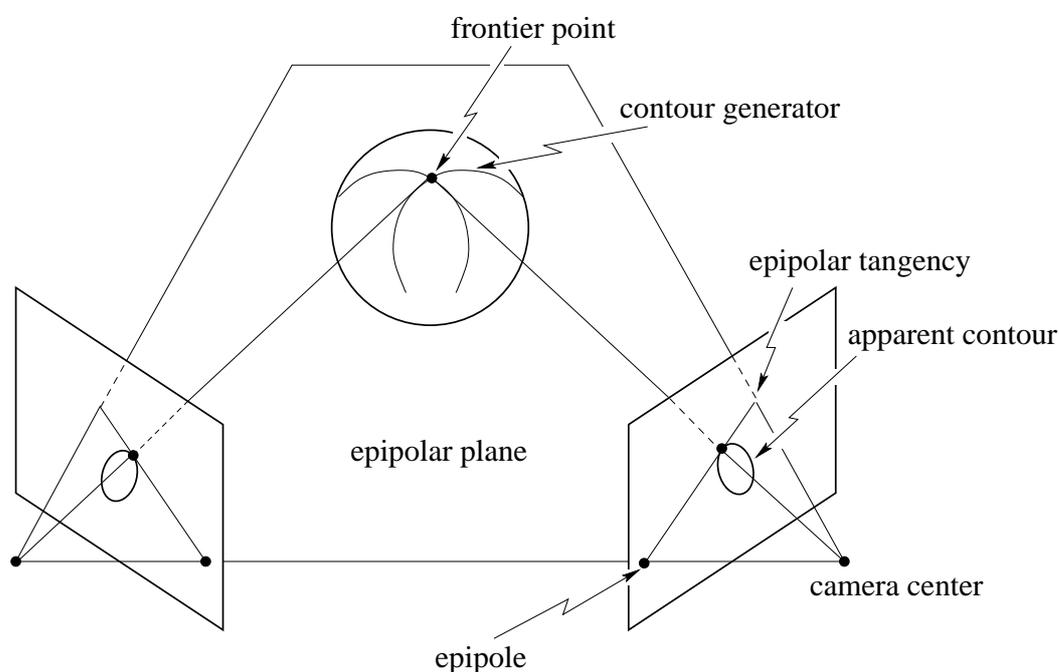


Figure 3.3. The frontier point is a fixed point on the surface, corresponding to the intersection of two contour generators. The epipolar lines corresponding to the frontier point are tangent to the profiles. Images courtesy of Kwan-Yee Kenneth Wong.

The problem in estimating the epipolar geometry from profiles lies in the difficulty of establishing image correspondences to be used in the computation of the cost function (3.24). The solution comes from property 1: the geometric error to be minimised will be the distance between epipolar lines and the closest tangent points on the profiles [25, 24, 28], as shown in figure 3.4.

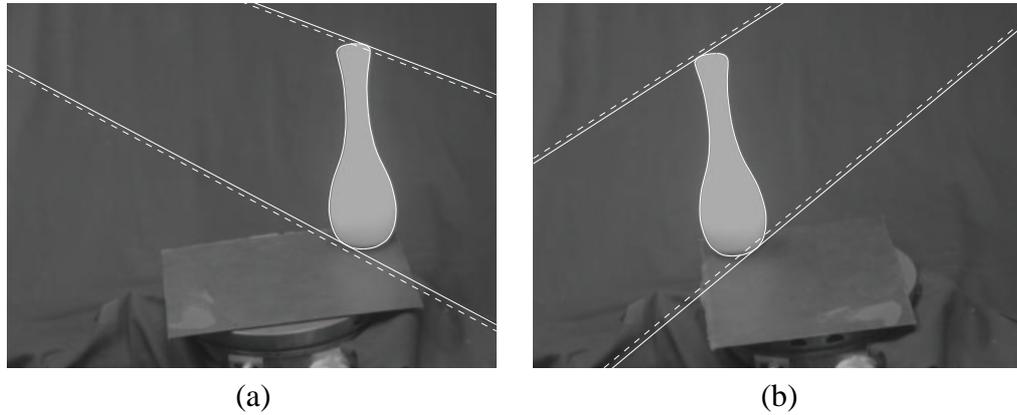


Figure 3.4. Geometric error for corresponding epipolar tangencies. The dashed line on each image is the epipolar line corresponding to the solid epipolar line on the other image. The distance between the dashed lines and the tangency point on the solid line is the geometric error to be used in the minimisation of (3.24).

The direct application of this method, however, is not practical. First, seven epipolar tangencies must be available, which is only possible when the object being viewed has a very rich geometry. Second, unless the initialisation of parameters in the optimisation process is made very close to the true solution, the algorithm will be poised to get trapped in a local minima. An example of this situation is shown in figure 3.5.

3.3.1 Affine Epipolar Geometry of Smooth Surfaces

When an affine approximation can be made, the application of property 1 can be greatly simplified [111]. Instead of searching for the full set of parameters of the fundamental matrix, one can initialise ψ and ψ' at arbitrary values. Then, epipolar lines tangent to the profiles at these angles can be easily obtained. Let \mathbf{l}_1 and \mathbf{l}_2 be two such lines, and let \mathbf{l}'_1 and \mathbf{l}'_2 be the corresponding epipolar lines at the second image. The distances from these lines to the origins of the respective image coor-

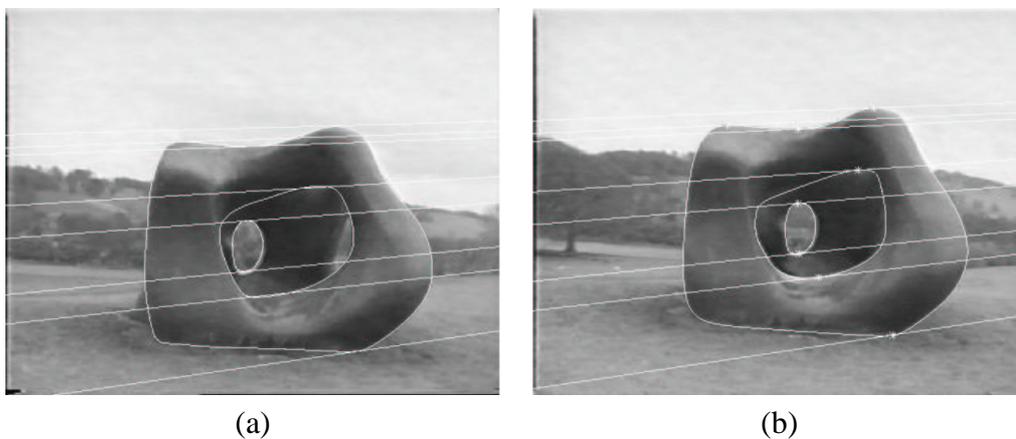


Figure 3.5. Result of a direct search for epipolar tangencies satisfying property 1. Although there is no ground truth available, a qualitative analysis shows that the solution found corresponds to a local minima: since the camera motion is roughly planar, one of the epipolar lines should be nearly parallel to the horizon [4], which does not happen. Images courtesy of Roberto Cipolla.

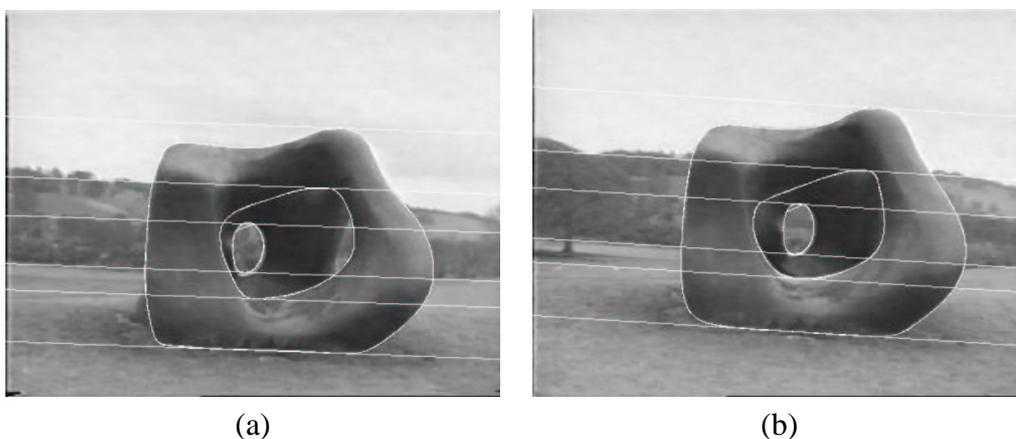


Figure 3.6. Estimation of epipolar geometry from profiles under affine approximation. The epipolar lines are parallel to the horizon, as expected from planar motion (compare with figure 3.5), and therefore the result is qualitatively correct.

dinate systems can be computed as d_i and d'_i , $i = 1, 2$. Finally, the solution of the

system of equations

$$\begin{aligned} d_2 &= \alpha d_1 - \beta \\ d'_2 &= \alpha d'_1 - \beta \end{aligned} \tag{3.35}$$

gives the parameters α and β . The remaining tangent lines can be used to compute a cost function similar to (3.24). The minimum number of epipolar tangencies necessary to solve the problem is now four, a more realistic figure. Moreover, the dimension of the search space has been reduced to two. This procedure is summarised in algorithm 3.1, and experimental results are shown in figure 3.6.

Algorithm 3.1 Estimation of the affine epipolar geometry from profiles.

initialise angles ψ and ψ' in (3.32);
while not converged **do**
 find tangents to the profiles at the angles ψ and ψ' ;
 select any two tangents \mathbf{l}_1 and \mathbf{l}_2 at the first image corresponding to tangents \mathbf{l}'_1 and \mathbf{l}'_2 at the second image;
 compute the distances d_i and d'_i , $i = 1, 2$ from the lines to the origins of the respective image coordinate systems;
 solve the equations (3.35);
 update ψ and ψ' to minimise (3.24) for the remaining epipolar lines;
end while

3.4 Geometry of Multiple Cameras

The fundamental matrix is the specialisation for two views of general *multicamera tensors* [64, 45, 71], which are algebraic expressions for the constraints the matching of features along images imposes on the projective basis of the scene and the cameras. Besides the fundamental matrix, the best known of these multicamera tensors are the *trifocal tensor* [136, 137, 64, 66, 45, 46, 47, 18] and the *quadfocal*

tensor [45, 19, 71, 138]. It has been shown in [45] that the quadrilinear relations are linear combinations of the bilinear ones (expressed by the fundamental matrices) and the trilinear ones (expressed by the trifocal tensor), and that any higher multilinear relation can be obtained from the bilinear, trilinear and quadrilinear ones. This section will briefly review some aspects of the trifocal tensor and generic multilinear relations.

3.4.1 Trifocal Tensors

This section makes extensive use of the Einstein notation for tensors [39] and definitions introduced in [66]. The matrix \mathbf{A} given by

$$\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1M} \\ \vdots & \ddots & \vdots \\ a_{N1} & \cdots & a_{NM} \end{bmatrix} \quad (3.36)$$

can be written as $\mathbf{A} = a_i^j$, where the subscript is the index of the column, and the superscript is the index of the row. Therefore, a_i represents a row vector and a^j represents a column vector. In general, $a_{i_1 i_2 \dots i_M}^{j_1 j_2 \dots j_N}$ represents a tensor of degree $N + M$. Finally, any equation of the type $\sum_{i=1}^N a_i b^i = 0$ is shortened to $a_i b^i = 0$, following the Einstein summation convention.

Consider the projective cameras $\mathbf{P} = [\mathbb{I} \quad \mathbf{0}] = \delta_i^j$, $\mathbf{P}' = a_i'^j$ and $\mathbf{P}'' = b_i''^j$, and let $\boldsymbol{\lambda} = \lambda_i$, $\boldsymbol{\lambda}' = \lambda_i'$ and $\boldsymbol{\lambda}'' = \lambda_i''$ be the homogeneous representation of the images at \mathbf{P} , \mathbf{P}' and \mathbf{P}'' of a line \mathbf{L} in space. Therefore, the planes $\pi_j = p_j^i \lambda_i$, $\pi_j' = p_j'^i \lambda_i'$ and $\pi_j'' = p_j''^i \lambda_i''$ intersect at \mathbf{L} . This condition can be used to derive the expression

$$\lambda_i = \lambda_j' \lambda_k'' T_i^{jk}, \quad (3.37)$$

where

$$T_i^{jk} = a_i^j b_4^k - a_4^j b_i^k \quad (3.38)$$

defines the *trifocal tensor* [136, 66]. Let now \mathbf{u} , \mathbf{u}' and \mathbf{u}'' be the images of a point in 3D space, each one taken from the camera with corresponding superscript (see figure 3.7). Then,

$$u''^l = u^k (u'^i T_k^{jl} - u'^j T_k^{il}). \quad (3.39)$$

The trifocal tensor is a $3 \times 3 \times 3$ tensor that plays for triplets of cameras the same role that the fundamental matrix plays for stereo systems. Observe that (3.37) and (3.39) are linear equations in the entries of the tensor that can be used to transfer points and lines matched in a pair of views to a third one. Since (3.37) is an equality in homogeneous coordinates, it provides three equations in the entries of T_i^{jk} , two of which are linearly independent. Analogously, (3.39) provides nine equations in the entries of T_i^{jk} , four of which are linearly independent [66]. The general expression of the *trilinear constraint* is

$$u^i \lambda'_j \lambda''_k T_i^{jk} = 0. \quad (3.40)$$

Parameterisation of the Trifocal Tensor. There are twenty-seven entries in the trifocal tensor. However, these entries are not independent. Consider three arbitrary camera matrices, which, together, hold $33 = 3 \times 11$ dof. The trifocal tensor does not depend on the underlying projective frame, i.e., the same trifocal tensor can be obtained by transforming the cameras according to an arbitrary projective transfor-

mation H_i^m , which has fifteen dof. Thus, the trifocal tensor has $18 = 33 - 15$ dof. The twenty-seven entries of the trifocal tensor are related through a set of nonlinear constraints that have been thoroughly investigated [46, 47, 18].

Minimal parameterisations of the trifocal tensor has been given in [148, 91, 47]. The parameterisation in [148] is given in terms of the invariants of six points in three views [127, 162], and it does not describe an explicit function that maps any given eighteen numbers into a trifocal tensor. This difficulty is overcome in [47], where the trifocal tensor is explicitly described in terms of eighteen parameters. However, a simpler parameterisation can be obtained, in which it is not necessary to solve any polynomial equation to obtain the tensor, therefore avoiding multiple solutions. Consider the transformation H_i^m given by

$$\mathbf{H} = \begin{bmatrix} w_1 \mathbb{I} & \mathbf{0} \\ \boldsymbol{\pi}^T & w_2 \end{bmatrix}, \quad (3.41)$$

where $w_i \in \mathbb{R} \setminus \{0\}$, $i = 1, 2$. Applying this transformation to the camera matrices \mathbf{P} , \mathbf{P}' and \mathbf{P}'' , the form of p_i^j is preserved for whichever values of d , $\boldsymbol{\pi}$ and s are used, which can then be chosen such that

$$\mathbf{P}' = \begin{bmatrix} 0 & 0 & 0 & 1 \\ a_1^2 & 1 & a_3^2 & a_4^2 \\ a_1^3 & a_2^3 & a_3^3 & a_4^3 \end{bmatrix} = [\mathbf{A} \ \mathbf{a}]. \quad (3.42)$$

Now there is no remaining degree of freedom in \mathbf{H} , but the overall scale of \mathbf{P}'' can

still be set so that

$$\mathbf{P}'' = \begin{bmatrix} b_1^1 & b_2^1 & b_3^1 & 1 \\ b_1^2 & b_2^2 & b_3^2 & b_4^2 \\ b_1^3 & b_2^3 & b_3^3 & b_4^3 \end{bmatrix} = [\mathbf{B} \mathbf{b}]. \quad (3.43)$$

Using (3.38) to compute T_i^{jk} one obtains an explicit minimal parameterisation of the trifocal tensor:

$$\mathbf{T}^k = \begin{bmatrix} -b_1^k & -b_2^k & -b_3^k \\ b_4^k a_1^2 - a_4^2 b_1^k & b_4^k - a_4^2 b_2^1 & b_4^k a_3^2 - a_4^2 b_3^k \\ b_4^k a_1^3 - b_4^3 b_1^k & b_4^k a_2^3 - a_4^3 b_2^k & b_4^k a_3^3 - a_4^3 b_3^k \end{bmatrix}, \quad (3.44)$$

where $b_4^1 = 1$. The parameterisation discussed in [47] is not general, since the entry 111 of the tensor is arbitrarily made equal to one, which may not always be possible, although the conditions under which this situation is verified are not known. The parameterisation given in (3.44) is also not general: in order to make \mathbf{P}' and \mathbf{P}'' as shown in (3.42) and (3.43), the left epipoles of the fundamental matrices obtained from \mathbf{P} and \mathbf{P}' and from \mathbf{P} and \mathbf{P}'' must have x -coordinate different from zero. However, this is a condition whose validity can be verified in advance, and an alternative parameterisation can be employed should it be found untrue.

Computation of the Trifocal Tensor. Given N_p triplets of points and N_l triplets of lines matched in three views such that $4N_p + 2N_l \geq 26$, (3.37) and (3.39) can be used to compute the entries of the trifocal tensor by a linear algorithm [66], albeit the solution will not satisfy the eight nonlinear constraints described in [18]. However, using the parameterisation in (3.44) it is possible to recover matrices \mathbf{P}'

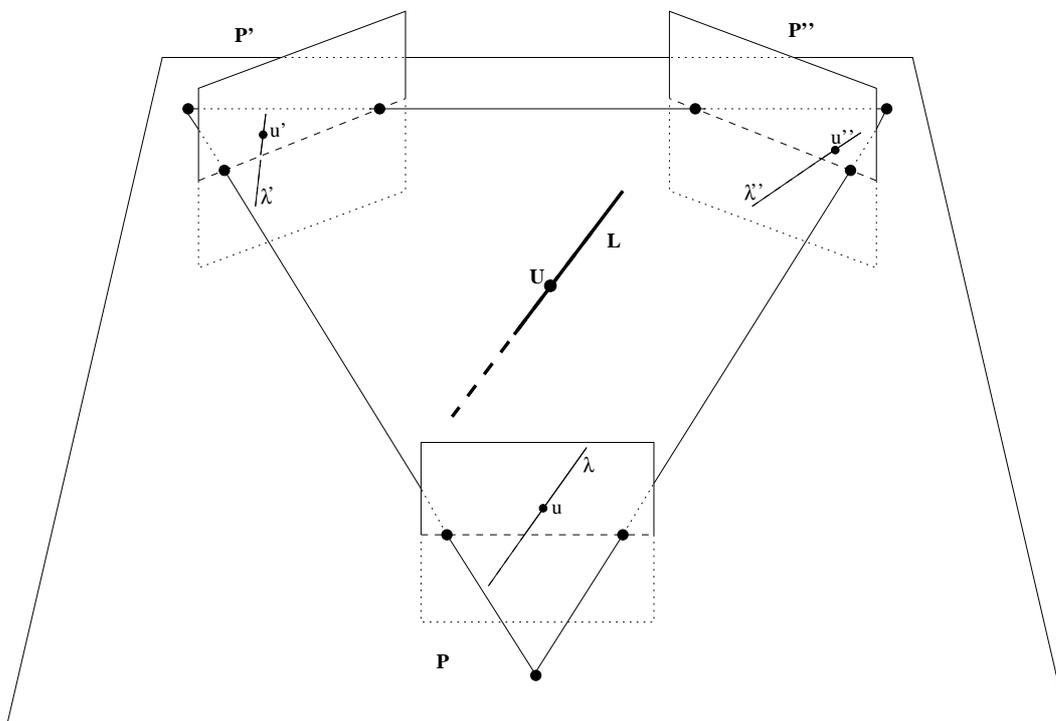


Figure 3.7. Geometric representation of the trilinear constraint.

and P'' as in (3.42) and (3.43) by solving an overconstrained set of linear equations. This is another advantage of using (3.44) over the parameterisation described in [47], which demands the iterative solution of a nonlinear set of equations to obtain the camera matrices and simultaneously impose the necessary constraints over the entries of the tensor computed from the linear algorithm.

To recover the matrices \mathbf{P}' and \mathbf{P}'' , one must first solve the linear system

$$\begin{bmatrix} 1 & 0 & -T_1^{11} & 0 & 0 & -T_1^{21} & 0 & 0 \\ 0 & 1 & -T_3^{11} & 0 & 0 & -T_3^{21} & 0 & 0 \\ 0 & 0 & -T_2^{11} & 0 & 0 & -T_2^{21} & 0 & 0 \\ 1 & 0 & 0 & -T_1^{12} & 0 & 0 & -T_1^{22} & 0 \\ 0 & 1 & 0 & -T_3^{12} & 0 & 0 & -T_3^{22} & 0 \\ 0 & 0 & 0 & -T_2^{12} & 0 & 0 & -T_2^{22} & 0 \\ 1 & 0 & 0 & 0 & -T_1^{13} & 0 & 0 & -T_1^{23} \\ 0 & 1 & 0 & 0 & -T_3^{13} & 0 & 0 & -T_3^{23} \\ 0 & 0 & 0 & 0 & -T_2^{13} & 0 & 0 & -T_2^{23} \end{bmatrix} \begin{bmatrix} a_1^2 \\ a_3^2 \\ a_4^2 \\ \frac{a_4^2}{b_4^2} \\ \frac{a_4^2}{b_4^3} \\ w \\ wb_4^2 \\ wb_4^3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -1 \\ 0 \\ 0 \\ -1 \\ 0 \\ 0 \\ -1 \end{bmatrix}, \quad (3.45)$$

where w is the scale factor of the tensor. Then, $\mathbf{B} = b_i^j$ can be computed as

$$b_i^j = -wT_i^{1j}. \quad (3.46)$$

The remaining parameters are a_i^3 , $i = 1, \dots, 4$, which can be obtained as the least squares solution of

$$\begin{bmatrix} \mathbb{I} & (T_i^{11})^T \\ b_4^2 \mathbb{I} & (T_i^{12})^T \\ b_4^3 \mathbb{I} & (T_i^{13})^T \end{bmatrix} \begin{bmatrix} (a_i^3)^T \\ a_4^3 \end{bmatrix} = \begin{bmatrix} (T_i^{31})^T \\ (T_i^{32})^T \\ (T_i^{33})^T \end{bmatrix}. \quad (3.47)$$

Once the matrices \mathbf{P}' and \mathbf{P}'' are found, an iterative procedure can be used to refine the eighteen parameters in (3.44) in order to minimise the geometric error in the transfer of points and lines.

Affine Trifocal Tensor. In the affine case (see section 2.3.2), the trifocal tensor must hold $12 = 3 \times 8 - 12$ dof, corresponding to the 3×8 dof of three affine cameras, minus the twelve dof of a general affine transformation. Therefore, by specialising the cameras \mathbf{P}' and \mathbf{P}'' to be affine, a simpler parameterisation of the trifocal tensor $T_i^{j,k}$ can be obtained, given by

$$\mathbf{T}^1 = \begin{bmatrix} -b_1^1 & -b_2^1 & -b_3^1 \\ a_1^2 - a_4^2 b_1^1 & 1 - a_4^2 b_2^1 & a_3^2 - a_4^2 b_3^1 \\ 0 & 0 & a_3^3 \end{bmatrix}, \quad (3.48)$$

$$\mathbf{T}^2 = \begin{bmatrix} -b_1^2 & -b_2^2 & -b_3^2 \\ b_4^2 a_1^2 - b_4^2 b_1^2 & b_4^2 - a_4^2 b_2^2 & b_4^2 a_3^2 - b_4^2 b_3^2 \\ 0 & 0 & b_4^2 a_3^3 - a_4^3 b_3^2 \end{bmatrix}, \quad (3.49)$$

$$\mathbf{T}^3 = \begin{bmatrix} 0 & 0 & -b_3^3 \\ 0 & 0 & -a_4^2 b_3^3 \\ 0 & 0 & -a_4^3 b_3^3 \end{bmatrix}. \quad (3.50)$$

Although in a different form, this is the same result obtained in [110, 46], but now the parameterisation is minimal, involving only twelve terms.

3.4.2 General Multilinear Relations

A natural extension of the trifocal tensor is the *quadfocal tensor* [45, 19, 71, 138], which is a $3 \times 3 \times 3 \times 3$ tensor encoding the geometric relations between four images. By considering the cameras \mathbf{P} , \mathbf{P}' , \mathbf{P}'' as before plus a new camera \mathbf{P}''' , the entries

$ijkl$ of a quadfocal tensor Q can be given by [71]

$$Q^{ijkl} = \det \begin{bmatrix} \mathbf{P}_i \\ \mathbf{P}'_j \\ \mathbf{P}''_k \\ \mathbf{P}'''_l \end{bmatrix}. \quad (3.51)$$

If the matrix \mathbf{P}''' is assumed to have the same format as the matrix \mathbf{P}'' in (3.43), a parameterisation of the quadfocal tensor with only $29 = 4 \times 11 - 15$ terms can be directly obtained from (3.51). This approach can be easily extended to parameterise tensors of any degree.

3.5 Conclusions

This chapter presented a summary of epipolar geometry and multicamera relations. It followed [164] in its approach to the introduction of the fundamental matrix, and discussed different parameterisations for \mathbf{F} . It also addressed some properties of the essential matrix, which are further developed in chapter 5 to produce a self-calibration algorithm. Some techniques for computing the fundamental matrix were also discussed in this chapter.

The review of the fundamental matrix was then specialised to affine cameras, and a novel parameterisation of the affine fundamental matrix was derived. This was followed by a brief overview of the connection between epipolar geometry and smooth surfaces, culminating in the development of a practical algorithm for estimating the affine epipolar geometry of a stereo rig from profiles. Experiments with real data were performed, and the results were compared to those of a well-

known method [24]. A qualitative analysis demonstrated the validity of the new approach.

The chapter was closed with a review of geometry of multicamera systems, both projective and affine. This review focused on the trifocal tensor, for which a novel minimal parameterisation was introduced. This parameterisation was shown to be easily extendable, providing minimal representations for higher degree multiview tensors.

Chapter 4

Epipolar Geometry from Profiles Under Circular Motion

Men go abroad to wonder [...] at the circular motions of the stars; and they pass by themselves without wondering.

St. Augustine (354–430), *Confessions*, book X.

4.1 Introduction

Methods for motion estimation and 3D reconstruction from point or line correspondences in a sequence of images have achieved a high level of sophistication, with impressive results [147, 84, 50]. Nevertheless, if corresponding points are not available the current techniques cannot be applied. That is exactly the case when the scene being viewed is composed of non-textured smooth surfaces, and in this situation the dominant feature in the image is the profile or apparent contour of the surface [85]. Besides, even when point correspondences can be established, the profile

still offers important clues for determining both motion and shape, and therefore it should be used whenever available.

This chapter addresses the problem of motion estimation and reconstruction of 3D models from profiles of an object rotating on a turntable, obtained from a fixed camera. Its main contribution is the development of a practical and accurate technique for solving this problem from profiles alone, which is accurate enough to allow the reconstruction of the object. No correspondence between points or lines are necessary, although the method proposed can be equally used when these features are available, without any further adaptation. Symmetry properties of the surface of revolution swept out by the rotating object are exploited to obtain the image of the rotation axis and the homography relating epipolar lines, in a robust and elegant way. These, together with geometric constraints for images of rotating objects, are then used to obtain first the image of the horizon, which is the projection of the plane that contains the camera centres, and then the epipoles, thus fully determining the epipolar geometry of the sequence of images. The estimation of the epipolar geometry by this sequential approach (image of rotation axis — homography — image of the horizon — epipoles) avoids many of the problems usually found in other algorithms for motion recovery from profiles. In particular, the search for the epipoles, by far the most critical step, is carried out as a simple one-dimensional optimisation problem. The initialisation of the parameters is trivial and completely automatic for all stages of the algorithm. After the estimation of the epipolar geometry, the Euclidean motion is recovered using the fixed intrinsic parameters of the camera, obtained either from a calibration grid or from self-calibration techniques. Finally, the spinning object is reconstructed from its profiles, using the motion estimated in the previous stage. Results from real data are presented, demonstrating

the efficiency and usefulness of the proposed method.

4.2 Previous Works and Bibliographic Review

The first attempts to approach the problem of motion estimation from apparent contours date back to Rieger, in 1986 [128], who introduced the concept of frontier point, interpreted as “centers of spin” [*sic*] of the image motion. That paper dealt with the case of frontoparallel orthographic projection, which is a rather restrictive situation. This idea was further developed by Porrill and Pollard [125], who recognised the frontier point as a fixed point on the surface, corresponding to the intersection of two consecutive contour generators [27] (see section 3.3). The connection between the epipolar geometry and the frontier points was established in [56], and an algorithm for motion estimation from profiles was introduced in [25]. Related works also include [6], where a technique based on registering the images using a planar curve was first developed. This method was implemented in [33], which also showed results of reconstruction from the estimated motion. In [111] the algorithm presented in [25] was specialised to the affine case. The work in [79] presents a method where the affine approximation is used to bootstrap the full projective case.

Initial steps towards a solution for the problem of reconstruction from apparent contours with known camera motion were given by Barrow and Tenenbaum, in 1981 [7], where a technique to compute surface normals was introduced. Koenderink [85, 87] established relations between the differential geometry of a surface and the differential geometry of its profiles. This work was extended in [55], where algorithms for computing the curvature of a surface from its profiles were developed and implemented for orthographic projection. In [158] a reconstruction

method based on parameterising the surface by *radial curves* was developed. Better results can be achieved by using an *epipolar parameterisation*, together with an interpolation using the *osculating circle*, as introduced in [27]. Further refinements were obtained in [14, 146], and a simple technique was developed in [161], based on a finite-difference implementation of [27]. Despite its simplicity, the method developed in [161] renders results comparable to those in [14] and [27], and was therefore the technique used here.

This work makes use of symmetry properties [165, 95, 167, 37] of the surface of revolution swept out by the rotating object to overcome the main difficulties and drawbacks present in other methods that have attempted to estimate motion from profiles, namely: the need for a very good initialisation for the epipolar geometry and an unrealistic demand for a large number of *epipolar tangencies* [25, 6, 5] (here as few as two epipolar tangencies are needed), restriction to linear motion [130] (whereas circular motion is a more practical situation), or the use of an affine approximation [111, 161] (which may be used only for shallow scenes, section 3.3.1). After obtaining the motion, the reconstruction can be achieved by a simple technique [161], based on the epipolar parameterisation [27], which extends the common triangulation methods from points to profiles.

An interesting comparison can be made between the work presented here and [50]. Both papers tackle the same problem, but while in [50] hundreds of points were tracked and matched for each pair of adjacent images, it will be shown here that a solution can be obtained even when only two epipolar tangencies are available, with at least comparable results.

This chapter begins by describing a method to obtain the image of the rotation axis and the coordinates of a special vanishing point, used in the parameterisation of

the fundamental matrix under circular motion, from symmetry properties of the profile of the surface of revolution swept out by an object placed on a turntable. These provide the homography component of the fundamental matrix in a plane plus parallax representation. The epipolar constraint is then used to estimate the epipoles for each pair of images in the sequence. These epipoles should be collinear, and the line containing them corresponds to the horizon. Due to noise, this alignment will not be verified, and a line is robustly fitted to the cloud of epipoles to provide an estimate for the horizon. Once this estimate is available, the epipolar constraint is again employed to recompute the epipoles with a minimal parameterisation specialised to circular motion [159]. The epipoles are now constrained to lie on the horizon, providing an accurate estimate for the epipolar geometry of each pair of images in the sequence. Intrinsic parameters, either computed from a self-calibration algorithm or precomputed by any standard calibration technique, can then be used together with the fundamental matrices to determine the camera motion.

Section 4.3 reviews the symmetry properties exhibited by the image of a surface of revolution summarised in the form of the *harmonic homology*. Section 4.4 establishes the relationship between this transformation and the epipolar geometry, and also presents two useful parameterisations of the fundamental matrix. These parameterisations allow the estimation of the epipoles to be carried out as independent one-dimensional searches, avoiding local minima and greatly reducing the computational complexity of the estimation. Section 4.5 presents the algorithm for motion recovery, and the implementation of the algorithm for real data is shown in section 4.6, which also makes comparisons with previous works. Experimental results using the estimated motion for reconstruction are shown in Section 4.7.

4.3 Symmetry in the Image of a Surface of Revolution

An object rotating about a fixed axis sweeps out a surface of revolution [54]. Symmetry properties of the image of this surface of revolution can be exploited to estimate the parameters of the motion of the object in a simple and elegant way, as will be shown next. In the definitions that follow, points and lines will be referred to by their representation as vectors in homogeneous coordinates, as usual.

A 2D homography that keeps the pencil of lines through a point \mathbf{v} and the set of points on a line \mathbf{l} fixed is called a *perspective collineation* with centre \mathbf{v} and axis \mathbf{l} . A *homology* is a perspective collineation whose centre and axis are not incident (otherwise the perspective collineation is called an *elation*). Let \mathbf{a} be a point mapped by a homology onto a point \mathbf{a}' . It is easy to show that the centre of the homology, \mathbf{v} , and the points \mathbf{a} and \mathbf{a}' are collinear. Let \mathbf{q}_a be the line passing through these points, and \mathbf{v}_a be the intersection of \mathbf{q}_a and the axis \mathbf{l} . If \mathbf{a} and \mathbf{a}' are harmonic conjugates with respect to \mathbf{v} and \mathbf{v}_a , i.e., their cross-ratio is one, the homology is said to be a *harmonic homology* (see details in [133, 30] and also figure 4.1(a)). The matrix \mathbf{W} representing a harmonic homology with centre \mathbf{v} and axis \mathbf{l} in homogeneous coordinates is given by

$$\mathbf{W} = \mathbb{I} - 2 \frac{\mathbf{v}\mathbf{l}^T}{\mathbf{v}^T\mathbf{l}}. \quad (4.1)$$

The profile of a surface of revolution exhibits a special symmetry property, which can be described by a harmonic homology [95]. The next theorem gives a formal definition for this property:

Theorem 1 *The profile of a surface of revolution S viewed by a pinhole camera is invariant to the harmonic homology with axis given by the image of the rotation axis of the surface of revolution and centre given by the image of the point at infinity in a direction orthogonal to a plane that contains the rotation axis and the camera centre .*

The following lemma will be used in the proof of theorem 1.

Lemma 1 *Let $\mathbf{T} : \Gamma' \mapsto \Gamma'$ be a harmonic homology with axis \mathbf{l}' and centre \mathbf{v}' on the plane Γ' , and let $\mathbf{H} : \Gamma' \mapsto \Gamma$ be a bijective 2D homography between the planes Γ' and Γ . Then, the transformation $\mathbf{W} = \mathbf{H}\mathbf{T}\mathbf{H}^{-1} : \Gamma \mapsto \Gamma$ is a harmonic homology with axis $\mathbf{l} = \mathbf{H}^{-\mathbf{T}}\mathbf{l}'$ and centre $\mathbf{v} = \mathbf{H}\mathbf{v}'$.*

Proof: Since \mathbf{H} is bijective, \mathbf{H}^{-1} exists. Then

$$\begin{aligned} \mathbf{W} &= \mathbf{H} \left(\mathbb{I} - 2 \frac{\mathbf{v}'\mathbf{l}'^{\mathbf{T}}}{\mathbf{v}'^{\mathbf{T}}\mathbf{l}'} \right) \mathbf{H}^{-1} \\ &= \mathbb{I} - 2 \frac{\mathbf{v}\mathbf{l}^{\mathbf{T}}}{\mathbf{v}^{\mathbf{T}}\mathbf{l}}, \end{aligned} \tag{4.2}$$

since $\mathbf{v}^{\mathbf{T}}\mathbf{l} = \mathbf{v}'^{\mathbf{T}}\mathbf{l}'$. □

The following corollary is a trivial consequence of lemma 1:

Corollary 1 *Let \mathbf{T} , \mathbf{H} , \mathbf{W} , Γ' and Γ be defined as in lemma 1. The transformation \mathbf{H} is an isomorphism between the structures (\mathbf{T}, Γ') and (\mathbf{W}, Γ) , i.e., $\forall \gamma \in \Gamma'$, $\mathbf{H}\mathbf{T}\gamma = \mathbf{W}\mathbf{H}\gamma$.*

An important consequence of lemma 1 and corollary 1 is that if a set of points $\hat{\zeta}$, e.g., the profile of a surface of revolution, is invariant to a harmonic homology \mathbf{T} , the

set ζ obtained by transforming $\hat{\zeta}$ by a 2D projective transformation \mathbf{H} is invariant to the harmonic homology $\mathbf{W} = \mathbf{H}\mathbf{T}\mathbf{H}^{-1}$.

Without loss of generality assume that the axis of rotation of the surface of revolution S is coincident with the y -axis of a right-handed orthogonal coordinate system. Considering a particular case of theorem 1 in which the pinhole camera $\hat{\mathbf{P}}$ is given by $\hat{\mathbf{P}} = [\mathbb{I} \ \mathbf{t}]$, where $\mathbf{t} = [0 \ 0 \ \alpha]^T$, for any $\alpha > 0$, symmetry considerations show that the profile $\hat{\zeta}$ of S will be bilaterally symmetric with respect to the image of the y -axis [116, 113], which corresponds to the line $\mathbf{q}_s = [1 \ 0 \ 0]^T$ in (homogeneous) image coordinates.

Proof of theorem 1 (particular case): Since $\hat{\zeta}$ is bilaterally symmetric about \mathbf{q}_s , there is a transformation \mathbf{T} that maps each point of $\hat{\zeta}$ onto its symmetric counterpart, given by

$$\mathbf{T} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (4.3)$$

However, as any bilateral symmetry transformation, \mathbf{T} is also a harmonic homology, with axis \mathbf{q}_s and centre $\mathbf{u}_x = [1 \ 0 \ 0]^T$, since

$$\mathbf{T} = \mathbb{I} - 2 \frac{\mathbf{u}_x \mathbf{q}_s^T}{\mathbf{u}_x^T \mathbf{q}_s}. \quad (4.4)$$

The transformation \mathbf{T} maps the set $\hat{\zeta}$ onto itself (although the points of $\hat{\zeta}$ are not mapped onto themselves by \mathbf{T} , but onto their symmetric counterparts), and thus $\hat{\zeta}$ is invariant to the harmonic homology \mathbf{T} . Since the camera centre lies on the z -axis of the coordinate system, the plane that contains the camera centre and the axis of

rotation is in fact the yz -plane, and the point at infinity orthogonal to this plane is $\mathbf{U}_x = [1 \ 0 \ 0 \ 0]^T$, whose image is \mathbf{u}_x . \square

Let \mathbf{P} be an arbitrary pinhole camera. The camera \mathbf{P} can be obtained by rotating $\hat{\mathbf{P}}$ about its optical centre by a rotation \mathbf{R} and transforming the image coordinate system of $\hat{\mathbf{P}}$ by introducing the intrinsic parameters represented by the calibration matrix \mathbf{K} . Let $\mathbf{KR} = \mathbf{H}$. Thus, $\mathbf{P} = \mathbf{H}[\mathbf{I} \ | \ \mathbf{t}]$, and the point \mathbf{U}_x in space with image \mathbf{u}_x in $\hat{\mathbf{P}}$ will project to a point $\mathbf{v}_x = \mathbf{H}\mathbf{u}_x$ in \mathbf{P} . Analogously, the line \mathbf{q}_s in $\hat{\mathbf{P}}$ will correspond to a line $\mathbf{l}_s = \mathbf{H}^{-T}\mathbf{q}_s$ in \mathbf{P} . It is now possible to derive the proof of theorem 1 in the general case.

Proof of theorem 1 (general case): Let ζ be the profile of the surface of revolution S obtained from the camera \mathbf{P} . Thus, the image of the bijection \mathbf{H} acting on the profile $\hat{\zeta}$ is ζ (or $\mathbf{H}\hat{\zeta} = \zeta$), and, using lemma 1, the transformation $\mathbf{W} = \mathbf{H}\mathbf{T}\mathbf{H}^{-1}$ is a harmonic homology with centre $\mathbf{v}_x = \mathbf{H}\mathbf{u}_x$ and axis $\mathbf{l}_s = \mathbf{H}^{-T}\mathbf{q}_s$. Moreover, from Corollary 1, $\mathbf{W}\mathbf{H}\hat{\zeta} = \mathbf{H}\mathbf{T}\hat{\zeta}$, or $\mathbf{W}\zeta = \mathbf{H}\mathbf{T}\hat{\zeta}$. From the particular case of the theorem 1 it is known that the profile $\hat{\zeta}$ will be invariant to the harmonic homology \mathbf{T} , so $\mathbf{W}\zeta = \mathbf{H}\hat{\zeta} = \zeta$. \square

When the camera is pointing directly towards the axis of rotation, the transformation that maps ζ onto its symmetric counterpart will be reduced to a skewed symmetry [81, 114, 20], which corresponds to a particular case of the harmonic homology in which the pole is at infinity. It is given by

$$\mathbf{S} = \frac{1}{\cos(\phi - \rho)} \begin{bmatrix} -\cos(\phi + \rho) & -2 \sin \rho \cos \phi & 2d \cos \phi \\ -2 \sin \phi \cos \rho & \cos(\phi + \rho) & 2d \sin \phi \\ 0 & 0 & \cos(\phi - \rho) \end{bmatrix}, \quad (4.5)$$

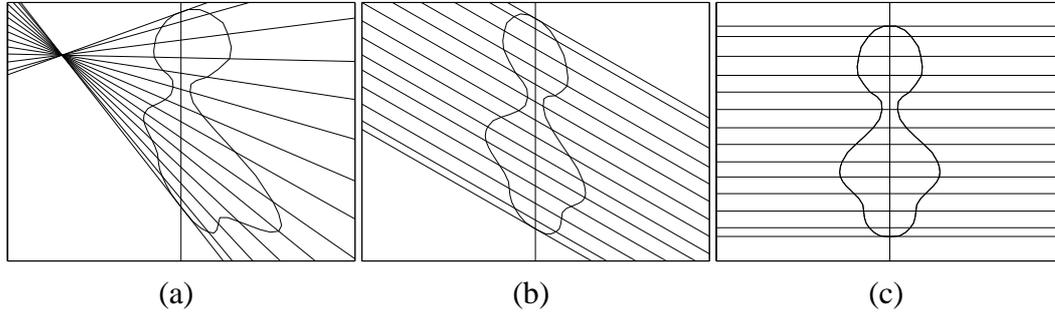


Figure 4.1. (a) Profile of a surface of revolution under general viewing conditions. The symmetry of the profile is represented by a harmonic homology defined by the image of the rotation axis and the pole. (b) When the camera is pointing towards the axis of rotation the transformation reduces to a skewed symmetry, which is a particular case of the harmonic homology with the pole at infinity. (c) If, additionally, the camera has zero skew and aspect ratio one, the transformation becomes a bilateral symmetry, in which the lines of symmetry are perpendicular to the image of the rotation axis.

where $\mathbf{l}_s = [\cos \rho \ \sin \rho \ -d]^T$ is the image of the rotation axis, with $d = u_0 \cos \rho + v_0 \sin \rho$. The angle ϕ gives the orientation of the *lines of symmetry*, which are the lines joining each point to its skew-symmetric counterpart (see figure 4.1(b)). The transformation \mathbf{S} has three dof.

If the camera also has zero skew and aspect ratio one, the transformation is further reduced to a bilateral symmetry, given by

$$\mathbf{B} = \begin{bmatrix} -\cos 2\rho & -\sin 2\rho & 2d \cos \rho \\ -\sin 2\rho & \cos 2\rho & 2d \sin \rho \\ 0 & 0 & 1 \end{bmatrix}. \quad (4.6)$$

The transformation now has only two dof, since the lines of symmetry are orthogonal to \mathbf{l}_s . A graphical representation of the bilateral symmetry, skewed symmetry and harmonic homology is shown in figure 4.1.

4.4 Parameterisations of the Fundamental Matrix

4.4.1 Fundamental Matrix under Circular Motion

The fundamental matrix corresponding to a pair of cameras related by a rotation around a fixed axis has a very special parameterisation, as shown in [159, 50], which can be expressed explicitly in terms of fixed image features under circular motion (image of rotation axis, pole and horizon, jointly holding 5 dof) and the relative angle of rotation (1 dof). A simpler derivation of this result will be shown here. Moreover, a novel parameterisation based on the harmonic homology will be introduced, providing a connection between the geometry of the complete sequence (harmonic homology) with the geometry of a single pair of images (fundamental matrix).

Consider the pair of camera matrices $\hat{\mathbf{P}}_1$ and $\hat{\mathbf{P}}_2$, given by

$$\begin{aligned}\hat{\mathbf{P}}_1 &= [\mathbb{I} \ \mathbf{t}] \\ \hat{\mathbf{P}}_2 &= [\mathbf{R}_y(\theta) \ \mathbf{t}],\end{aligned}\tag{4.7}$$

where

$$\mathbf{t} = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T\tag{4.8}$$

and

$$\mathbf{R}_y(\theta) = \begin{bmatrix} \cos \theta & 0 & \sin \theta \\ 0 & 1 & 0 \\ -\sin \theta & 0 & \cos \theta \end{bmatrix},\tag{4.9}$$

for $\theta \neq 0$. Let $\hat{\mathbf{F}}$ be the fundamental matrix relating $\hat{\mathbf{P}}_1$ and $\hat{\mathbf{P}}_2$. From (4.7), (4.8) and (4.9), it is easy to see that

$$\begin{aligned} \hat{\mathbf{F}} &= \begin{bmatrix} 0 & \cos \theta - 1 & 0 \\ \cos \theta - 1 & 0 & \sin \theta \\ 0 & -\sin \theta & 0 \end{bmatrix} \\ &= -\sin \theta \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}_{\times} + (\cos \theta - 1) \left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \right). \end{aligned} \quad (4.10)$$

Let \mathbf{U}_x , \mathbf{U}_y and \mathbf{U}_z be the points at infinity in the x , y and z directions, respectively, in world coordinates. Projecting these points using the camera $\hat{\mathbf{P}}_1$, we obtain the vanishing points \mathbf{u}_x , \mathbf{u}_y and \mathbf{u}_z given by

$$\mathbf{u}_x = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{u}_y = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{u}_z = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}. \quad (4.11)$$

The image of the horizon is the line \mathbf{q}_h , and the image of the rotation axis is the line \mathbf{q}_s , where

$$\mathbf{q}_s = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{q}_h = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}. \quad (4.12)$$

Substituting (4.11) and (4.12) into (4.10), the desired parameterisation is obtained:

$$\hat{\mathbf{F}} = -\sin \theta \left[[\mathbf{u}_x]_{\times} + \tan \frac{\theta}{2} (\mathbf{q}_s \mathbf{q}_h^T + \mathbf{q}_h \mathbf{q}_s^T) \right]. \quad (4.13)$$

The factor “ $-\sin \theta$ ” can be eliminated since the fundamental matrix is defined only up to an arbitrary scale. Assume now that the cameras $\hat{\mathbf{P}}_1$ and $\hat{\mathbf{P}}_2$ are transformed by a rotation \mathbf{R} about their optical centers and the introduction of a set of intrinsic parameters represented by the calibration matrix \mathbf{K} . The new pair of cameras, \mathbf{P}_1 and \mathbf{P}_2 , is related to $\hat{\mathbf{P}}_1$ and $\hat{\mathbf{P}}_2$ by

$$\begin{aligned} \mathbf{P}_1 &= \mathbf{H} \hat{\mathbf{P}}_1 \text{ and} \\ \mathbf{P}_2 &= \mathbf{H} \hat{\mathbf{P}}_2, \end{aligned} \quad (4.14)$$

where $\mathbf{H} = \mathbf{KR}$. The fundamental matrix \mathbf{F} of the new pair of cameras \mathbf{P}_1 and \mathbf{P}_2 is given by

$$\begin{aligned} \mathbf{F} &= \mathbf{H}^{-T} \hat{\mathbf{F}} \mathbf{H}^{-1} \\ &= \det(\mathbf{H}) [\mathbf{v}_x]_{\times} + \tan \frac{\theta}{2} (\mathbf{l}_s \mathbf{l}_h^T + \mathbf{l}_h \mathbf{l}_s^T), \end{aligned} \quad (4.15)$$

where $\mathbf{v}_x = \mathbf{H} \mathbf{u}_x$, $\mathbf{l}_h = \mathbf{H}^{-T} \mathbf{q}_h$ and $\mathbf{l}_s = \mathbf{H}^{-T} \mathbf{q}_s$. Since the fundamental matrix is defined only up to a scale factor, (4.15) can be rewritten as

$$\mathbf{F}(\theta) = [\mathbf{v}_x]_{\times} + \kappa \tan \frac{\theta}{2} (\mathbf{l}_s \mathbf{l}_h^T + \mathbf{l}_h \mathbf{l}_s^T), \quad (4.16)$$

where $\kappa = 1 / \det(\mathbf{H})$. The notation $\mathbf{F}(\theta)$ was used in (4.16) to emphasise that, for a given circular motion sequence, the parameters \mathbf{v}_x , \mathbf{l}_s , \mathbf{l}_h and κ are fixed, and the fundamental matrices associated with any pair of cameras in the sequence differs

only in the value of θ .

4.4.2 Parameterisation via Planar Harmonic Homology

The images of a rotating object are the same as the images of a fixed object taken by a camera rotating around the same axis, or by multiple cameras along that circular trajectory. Consider any two such cameras, denoted by \mathbf{P}_1 and \mathbf{P}_2 . If \mathbf{P}_1 and \mathbf{P}_2 point towards the axis of rotation and have zero skew and aspect ratio 1, their epipoles \mathbf{e}_1 and \mathbf{e}_2 will be symmetric with respect to the image of the rotation axis, or $\mathbf{e}_2 = \mathbf{T}\mathbf{e}_1$, according to figure 4.2. In a general situation, the epipoles will simply be related by the transformation $\mathbf{e}_2 = \mathbf{W}\mathbf{e}_1$. It is then straightforward to show that the corresponding epipolar lines \mathbf{l}_1 and \mathbf{l}_2 are related by $\mathbf{l}_2 = \mathbf{W}^{-T}\mathbf{l}_1$. This means that the pair of epipoles can be represented with only two parameters once \mathbf{W} is known. From (4.1) it can be seen that \mathbf{W} has only four dof.

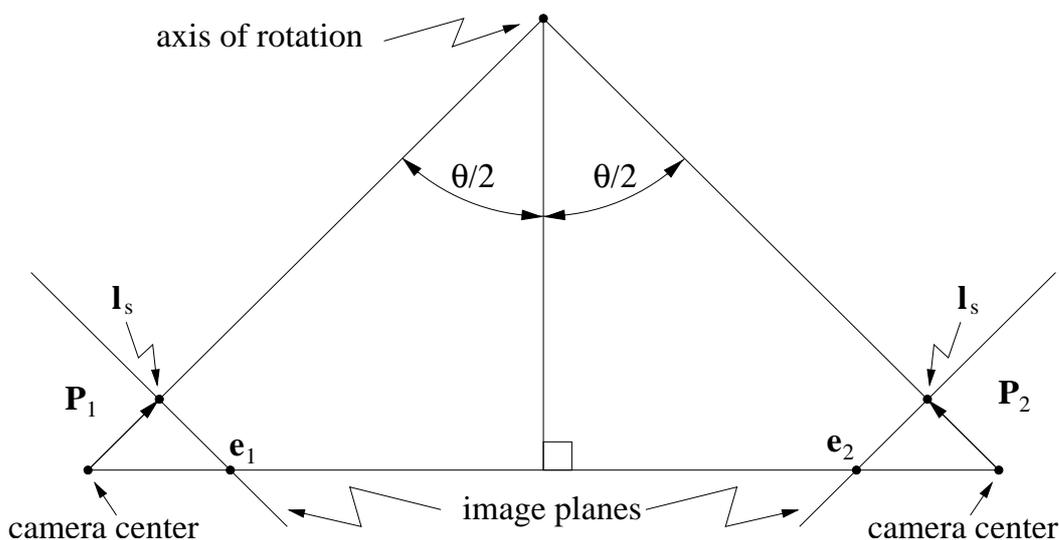


Figure 4.2. If the cameras are pointing towards the axis of rotation and their skew is zero and aspect ratio is 1, the epipoles \mathbf{e}_1 and \mathbf{e}_2 are symmetric with respect to the image of the rotation axis.

It is possible to express \mathbf{F} in (4.16) according to a planar plus parallax representation (see section 3.1.2) such that $\tilde{\mathbf{F}} = [\tilde{\mathbf{e}}']_{\times} \tilde{\mathbf{M}}$, where $\tilde{\mathbf{M}}^{-\text{T}}$ is any matrix that maps the epipolar lines from one image to the other, and $\tilde{\mathbf{e}}'$ is the epipole in the second image. From this discussion and section 4.3, it follows that

$$\mathbf{F} = [\mathbf{e}_2]_{\times} \mathbf{W}, \quad (4.17)$$

where, from (4.16), $\mathbf{e}_2 = \mathbf{v}_x - \kappa \tan \frac{\theta}{2} [\mathbf{I}_s]_{\times} \mathbf{I}_h$. Therefore, \mathbf{F} has only six dof: four to determine \mathbf{W} and two to fix \mathbf{e}_2 , in agreement with [159]. Note that in the case of skewed symmetry and bilateral symmetry, the dof of the fundamental matrix will be reduced to five and four respectively, corresponding to the decrease in the dof of the symmetry transformation. A full account of the dof of the fundamental matrix under different configurations is given in table 4.1.

From (4.17) it can be seen that the transformation \mathbf{W} corresponds to a plane induced homography. This means that the registration of the images can be done by using \mathbf{W} instead of a planar contour as proposed in [6, 33]. It has been discussed in section 3.1.2 that different choices of the plane that induces the homography in a plane plus parallax parameterisation of the fundamental matrix, such as the one in (4.17), will result in different homographies, although they will all generate the same fundamental matrix, since

$$\mathbf{F} = [\mathbf{e}_2]_{\times} \mathbf{W} = [\mathbf{e}_2]_{\times} [\mathbf{W} + \mathbf{e}_2 \mathbf{b}^{\text{T}}] \quad \forall \mathbf{b} \in \mathbb{R}^3. \quad (4.18)$$

The three parameter family of homographies $[\mathbf{W} + \mathbf{e}_2 \mathbf{b}^{\text{T}}]$ parameterised in \mathbf{b} has a one-to-one correspondence with the set of planes in \mathbb{R}^3 . The particular plane that induces the planar homology \mathbf{W} is given in the next theorem:

Theorem 2 *The planar homology \mathbf{W} relating the cameras \mathbf{P}_1 and \mathbf{P}_2 is induced by the plane Ξ that contains the axis of rotation and bisects the segment joining the optical centres of the cameras.*

Proof: The existence and uniqueness of Ξ satisfying the hypothesis of the theorem are trivial. Let $\mathbf{x}_1 = [1 \ 0 \ 0]^T$, $\mathbf{x}_2 = [0 \ 1 \ 0]^T$, and $\mathbf{x}_3 = [0 \ 0 \ 1]^T$. Without loss of generality, let

$$\begin{aligned}\mathbf{P}_1 &= \mathbf{KR}[\mathbb{I} \ \mathbf{x}_3] \quad \text{and} \\ \mathbf{P}_2 &= \mathbf{KR}[\mathbf{R}_y^\theta \ \mathbf{x}_3],\end{aligned}\tag{4.19}$$

where \mathbf{K} is the matrix of intrinsic parameters of \mathbf{P}_1 and \mathbf{P}_2 , \mathbf{R} is the rotation matrix relating the orientation of the coordinate system of \mathbf{P}_1 to the world coordinate system, and \mathbf{R}_y^θ is a rotation by θ about the y -axis of the world coordinate system, i.e.,

$$\mathbf{R}_y^\theta = \begin{bmatrix} \cos \theta & 0 & \sin \theta \\ 0 & 1 & 0 \\ -\sin \theta & 0 & \cos \theta \end{bmatrix}.\tag{4.20}$$

Therefore, $\forall \alpha, \beta \in \mathbb{R}$, the point $\mathbf{X} = [-\alpha \sin(\theta/2) \ \beta \ \alpha \cos(\theta/2)]^T$ lies on Ξ . Projecting \mathbf{X} using \mathbf{P}_1 and \mathbf{P}_2 , one obtains $\mathbf{u}_1 = \mathbf{KR}(\mathbf{X} + \mathbf{x}_3)$ and $\mathbf{u}_2 = \mathbf{KR}(\mathbf{R}_y^\theta \mathbf{X} + \mathbf{x}_3)$. Since

$$\mathbf{R}_y^\theta \mathbf{X} = \begin{bmatrix} \alpha \sin \theta \cos(\theta/2) - \alpha \cos \theta \sin(\theta/2) \\ \beta \\ \alpha \sin \theta \sin(\theta/2) + \alpha \cos \theta \cos(\theta/2) \end{bmatrix}$$

$$= \begin{bmatrix} \alpha \sin(\theta/2) \\ \beta \\ \alpha \cos(\theta/2) \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{X}, \quad (4.21)$$

or $\mathbf{R}_y^\theta \mathbf{X} = (\mathbb{I} - 2\mathbf{x}_1\mathbf{x}_1^T)\mathbf{X}$, we have $\mathbf{u}_2 = \mathbf{KR}[(\mathbb{I} - 2\mathbf{x}_1\mathbf{x}_1^T)\mathbf{X} + \mathbf{x}_3]$, or $\mathbf{u}_2 = (\mathbb{I} - 2\mathbf{KR}\mathbf{x}_1\mathbf{x}_1^T\mathbf{R}^{-1}\mathbf{K}^{-1})\mathbf{u}_1$. It can be shown [113] that $\mathbf{KR}\mathbf{x}_1 = \mathbf{v}_x$ and $\mathbf{x}_1^T\mathbf{R}^{-1}\mathbf{K}^{-1} = \mathbf{l}_s^T$, and thus $\mathbf{u}_2 = \mathbf{W}\mathbf{u}_1$. \square

A graphical representation of the result in theorem 2 is shown in figure 4.3.

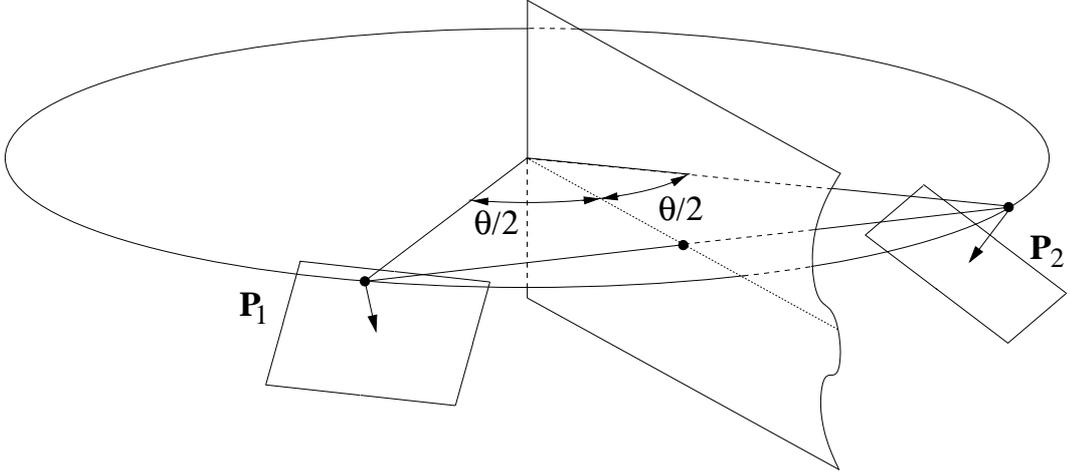


Figure 4.3. The harmonic homology is a homography induced by the plane that contains the axis of rotation and bisects the segment joining the camera centers.

4.5 Motion Estimation

Consider an object that undergoes a full rotation around a fixed axis. The envelope ε of its profiles is found by overlapping the images of the sequence and applying a Canny edge detector [17] to the resultant image. This envelope corresponds to the image of a surface of revolution, and thus it is harmonically symmetric. The

| Configuration | Parameterisation | dof |
|--|--------------------------------------|-------|
| General motion | $[\mathbf{e}]_{\times} \mathbf{M}$ | 2 + 5 |
| Circular motion | $[\mathbf{e}]_{\times} \mathbf{W}$ | 2 + 4 |
| Circular motion with camera pointing at axis of rotation | $[\mathbf{e}]_{\times} \mathbf{S}$ | 2 + 3 |
| Circular motion with camera pointing at axis of rotation and having zero skew and aspect ratio 1 | $[\mathbf{e}]_{\times} \mathbf{B}$ | 2 + 2 |
| Circular motion with camera pointing at axis of rotation and having zero skew and no rotation about the optical axis | $[\mathbf{e}]_{\times} \mathbf{B}_0$ | 2 + 1 |
| Pure translation | $[\mathbf{e}]_{\times}$ | 2 |
| Pure translation orthogonal to optical axis | $[\mathbf{e}_{\infty}]_{\times}$ | 1 |

Table 4.1. Analysis of the dof of the fundamental matrix for different types of motion with fixed intrinsic parameters.

homography \mathbf{W} related to ε is then found by sampling N points \mathbf{x}_i along ε and optimising the cost function

$$f_{\mathbf{W}}(\mathbf{v}_x, \mathbf{l}_s) = \sum_{i=1}^N \text{dist}(\varepsilon, \mathbf{W}(\mathbf{v}_x, \mathbf{l}_s)\mathbf{x}_i)^2, \quad (4.22)$$

where $\text{dist}(\varepsilon, \mathbf{W}(\mathbf{v}_x, \mathbf{l}_s)\mathbf{x}_i)$ is the orthogonal distance between the curve ε and the transformed sample point $\mathbf{x}'_i = \mathbf{W}(\mathbf{v}_x, \mathbf{l}_s)\mathbf{x}_i$. The estimation of \mathbf{W} is summarised in procedure 4.2.

The initialisation of the line \mathbf{l}_s and the point \mathbf{v}_x can be made very close to the global minimum by automatically locating one or more pairs of corresponding bitangents on the envelope. Given two bitangents $\mathbf{l}(\mathbf{p}_1, \mathbf{p}_2)$ and $\mathbf{l}(\mathbf{q}_1, \mathbf{q}_2)$ on the two sides of the profile ε with bitangent points $\mathbf{p}_1, \mathbf{p}_2$ and $\mathbf{q}_1, \mathbf{q}_2$, respectively (see figure 4.4), the intersection of the two bitangents ($\mathbf{l}(\mathbf{p}_1, \mathbf{p}_2), \mathbf{l}(\mathbf{q}_1, \mathbf{q}_2)$) and the intersection of the diagonals ($\mathbf{l}(\mathbf{p}_1, \mathbf{q}_2), \mathbf{l}(\mathbf{q}_1, \mathbf{p}_2)$) give two points defining a line that can be used as an estimate of \mathbf{l}_s . An estimate for the vanishing point \mathbf{v}_x is given by

the point of intersection of the lines $\mathbf{l}(\mathbf{p}_1, \mathbf{q}_1)$ and $\mathbf{l}(\mathbf{p}_2, \mathbf{q}_2)$. The initialisation of \mathbf{l}_s and \mathbf{v}_x from bitangents often provides an excellent initial guess for the optimisation problem. This is generally good enough to avoid any local minimum and allows convergence to the global minimum in a small number of iterations.

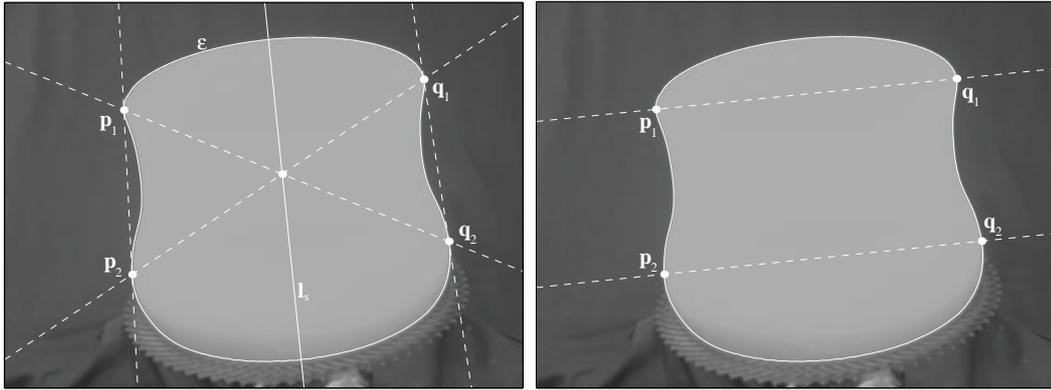


Figure 4.4. Initialisation of the optimisation parameters \mathbf{l}_s and \mathbf{v}_x from the bitangents and lines formed by bitangent points.

Procedure 4.2 *Get_Homology*: Estimation of the harmonic homology \mathbf{W} .

overlap the images in the sequence;
 extract the envelope ε of the profiles using a Canny edge detector;
 sample N points \mathbf{x}_i along ε ;
 initialise the axis of symmetry \mathbf{l}_s and the vanishing point \mathbf{v}_x using bitangents;
while not converged **do**
 compute the points $\mathbf{x}'_i = \mathbf{W}\mathbf{x}_i$;
 compute the distances between ε and \mathbf{x}'_i using (4.22);
 update \mathbf{l}_s and \mathbf{v}_x to minimise (4.22);
end while

After obtaining a good estimate of \mathbf{W} , one can then search for *epipolar tangencies* between pairs of images in the sequence using the parameterisation given by (4.17). To obtain a pair of corresponding epipolar tangents in two images, it is necessary to find a line tangent to one profile that is transformed by \mathbf{W}^{-T} to a line tangent to the profile in the other image (see figure 4.5). The search for correspond-

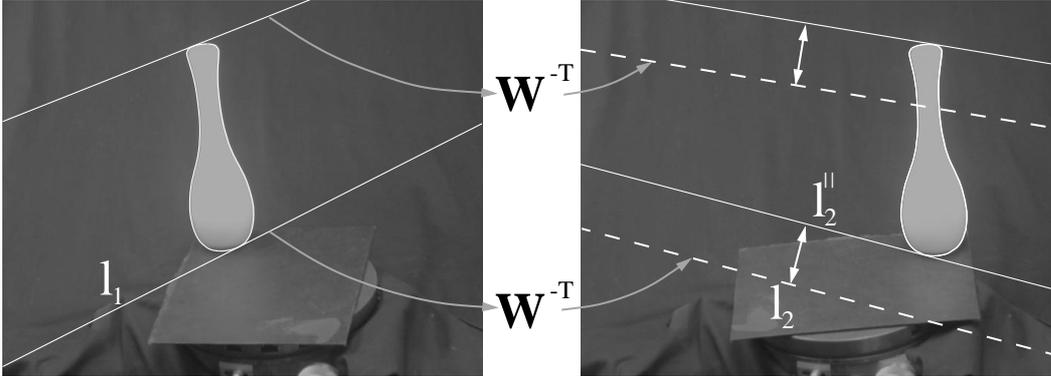


Figure 4.5. The line \mathbf{l}_1 tangent to the bottom of the profile in the first image is transferred to the line \mathbf{l}_2 in the second image by the harmonic homology. A line \mathbf{l}_2^{\parallel} parallel to \mathbf{l}_2 and tangent to the bottom of the profile is located and the distance between \mathbf{l}_2 and \mathbf{l}_2^{\parallel} drives the search for the orientation of \mathbf{l}_1 , which upon convergence will correspond to an epipolar tangent. An epipolar tangent at the top of the profile is obtained in the same way.

ing tangents may be carried out as a one-dimensional optimisation problem. The single parameter is the angle δ that defines the orientation of the epipolar line \mathbf{l}_1 in the first image, and the cost function is given by

$$f_{l_1}(\delta) = \text{dist}(\mathbf{W}^{-T}\mathbf{l}_1(\delta), \mathbf{l}_2^{\parallel}(\delta)), \quad (4.23)$$

where the function $\text{dist}(\mathbf{W}^{-T}\mathbf{l}_1(\delta), \mathbf{l}_2^{\parallel}(\delta))$ gives the distance between the transferred line $\mathbf{l}_2 = \mathbf{W}^{-T}\mathbf{l}_1$ and a line \mathbf{l}_2^{\parallel} parallel to \mathbf{l}_2 and tangent to the profile in the second image. Typical values of δ lie between -0.5 rad and 0.5 rad, or -30° and 30° . The shape of the cost function (4.23) for the profiles in figure 4.5 can be seen in figure 4.6.

The epipoles can then be computed as the intersection of epipolar lines in the same image. After obtaining this first estimate for the epipoles, the image of the horizon can then be found by robustly fitting a line \mathbf{l}_h to the initial set of epipoles, such that $\mathbf{l}_h^T \mathbf{v}_x = 0$. Figure 4.10 shows a typical output of procedure 4.3, together

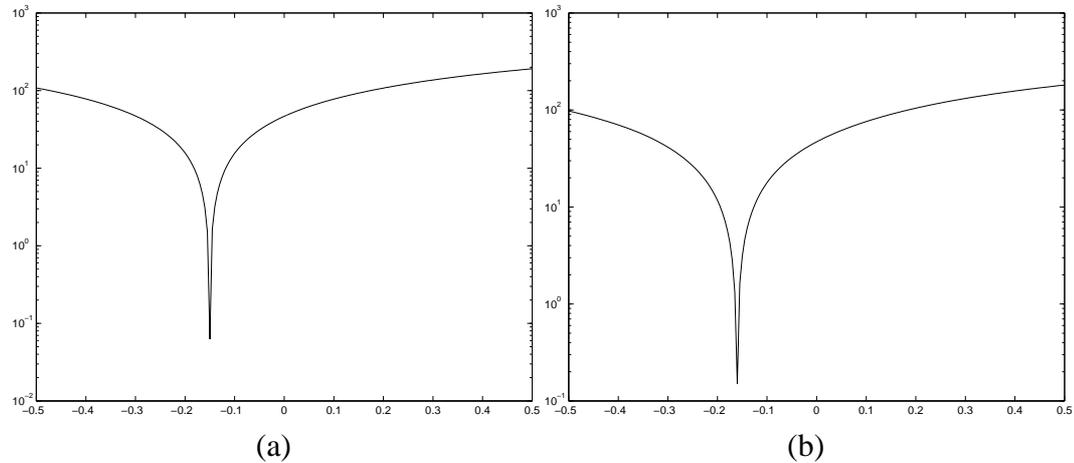


Figure 4.6. Plot of the cost function (4.23) for corresponding epipolar tangents near the top (a) and bottom (b) of the profiles in figure 4.5.

Procedure 4.3 *Get_Horizon*: Estimation of the horizon.

```

extract the profiles of the images using a Canny edge detector;
fit B-splines to the top and the bottom of the profiles;
for each selected pair of images do
  for top and bottom of profiles do
    initialise the angle  $\delta$  defining the orientation of the epipolar line at the first
    image;
    while not converged do
      find  $\mathbf{l}_1$  (see figure 4.5);
      compute the line  $\mathbf{l}_2 = \mathbf{W}^{-T}\mathbf{l}_1$ ;
      find the line  $\mathbf{l}_2^{\parallel}$  (see figure 4.5);
      compute the distance between  $\mathbf{l}_2$  and  $\mathbf{l}_2^{\parallel}$  using (4.23);
      update  $\delta$  to minimise (4.23);
    end while
  end for
  compute epipoles by intersecting epipolar lines tangent to the top and bottom
  of the profiles in each image;
end for
fit the horizon  $\mathbf{l}_h$  to the cloud of epipoles.

```

with the horizon \mathbf{l}_h fitted to the epipoles.

An alternative method to compute the epipoles is to register the profiles using the homology \mathbf{W} , eliminating the effects of rotation on the images, and then apply

any of the methods in [6, 130, 33], in a plane plus parallax approach. However, no advantage has been obtained by doing so, since to use this method it is necessary to find a common tangent between two profiles, which involves a search at least as complex as the one in procedure 4.3.

4.5.1 Estimation of the Epipoles Constrained to the Horizon

After estimating the horizon, the only missing term in the parameterisation of the fundamental matrix shown in (4.16) is the scale factor

$$\lambda = \kappa \tan \theta / 2. \quad (4.24)$$

This parameter can be found, again, by a one-dimensional search that minimizes the geometric error of transferred epipolar lines as shown in figure 4.7. Therefore, two distinct parameterizations of the fundamental matrix are used: (4.17) to obtain the cloud of epipoles and the horizon, and (4.16) to recompute the position of the epipoles constrained to lie on the horizon.

Procedure 4.4 *Get_Epipoles*: Estimation of the Epipoles.

```

for each selected pair of images do
  initialise the scale factor  $\lambda$  in (4.24);
  while not converged do
    compute a putative fundamental matrix using (4.16);
    locate epipolar tangents at the top and the bottom of profiles in both images;
    transfer epipolar tangents from the first image to the second the image using W;
    compute the geometric error as the distance between the lines transferred
    from the first image to the epipolar tangencies in the second image, as shown
    in figure 4.7;
    update  $\lambda$  to minimise the geometric error;
  end while
end for

```

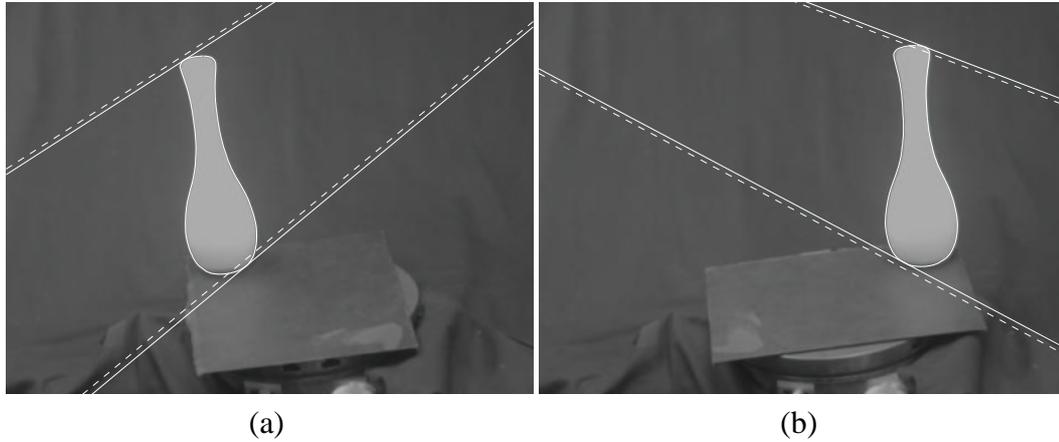


Figure 4.7. Once the horizon is computed, the location of the epipoles along this line can be refined by using (4.16). This figure shows the geometric error for transferred epipolar lines. The terms \mathbf{v}_x , \mathbf{l}_s and \mathbf{l}_h were obtained from procedure 4.2 and procedure 4.3. The solid lines in each correspond to tangents to the profile passing through the putative epipoles, and the dashed lines correspond to lines transferred from one image to the other by applying the harmonic homology \mathbf{W} . The sum of the distances between transferred lines and the corresponding tangent points is the geometric error that drives the search for the scale factor $\lambda = \kappa \tan \theta/2$ in (4.16). This scale factor was set to 100 in the figure, for better visualisation.

The overall procedure for estimating the epipolar geometry of a turntable sequence is shown in algorithm 4.1.

Algorithm 4.1 Estimation of the Epipolar Geometry.

- estimate the harmonic homology \mathbf{W} using `Get_Homology` (see procedure 4.2);
 - estimate the horizon \mathbf{l}_h using `Get_Horizon` (see procedure 4.3);
 - estimate the epipoles using `Get_Epipoles` (see procedure 4.4);
 - compute the fundamental matrices using (4.17);
-

4.5.2 Limitations of the Algorithm

There are some limitations on the applicability of the algorithms presented here:

Density of the sequence of images. If the number of images in the sequence is too small, or the angle of rotation between successive snapshots is too large, the envelope of the profiles no longer approximates the profile of a surface of revolution, and, therefore, procedure 4.2 will fail to correctly estimate the image of the rotation axis and the pole. In practice, this problem does not arise if the angles of rotation in a closed sequence are below 15° . This problem can be overcome by performing a simultaneous search for the harmonic homology and the rotation angles, at the expense of increasing the number of search parameters and therefore the complexity of the optimisation.

Symmetry of the object. If the object placed on the turntable is rotationally symmetric and its axis of symmetry coincides with the axis of rotation of the turntable, procedure 4.3 will fail. To understand this problem, consider the alternative formulation of procedure 4.3 in which the epipoles are computed by first registering the images by using the harmonic homology and then computing the epipoles as the intersection of common tangents to the profiles. Under the conditions described above, the registration of the profiles will not produce any effect, since the image of a surface of revolution with the same rotation axis as the turntable is invariant to the harmonic homology. Moreover, the profiles will coincide, and any tangent to one of the profiles will be a common tangent to the pair of profiles. Therefore, the position of the epipole will be undetermined. To avoid this problem it is enough to reposition the symmetric object over the turntable so that its symmetry axis does not coincide anymore with the turntable axis. The further the two axes are, the better. Of course, the placement of the object must not be so distant from the centre of the turntable as to remove it from the field of view. In the experiments shown in this paper using a vase and a head model, which are nearly rotationally symmetric in the

regions of interest (the top and the bottom of the objects), it was verified that the problem disappears if the axes are separated by a distance of about 50 pixels.

4.6 Implementation and Experimental Results

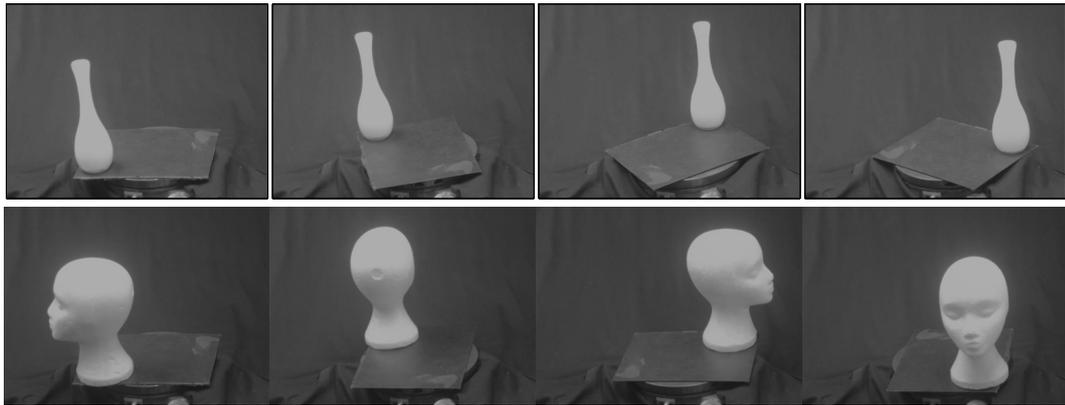


Figure 4.8. Top row shows four images of the vase. Bottom row shows four images of the head model.

The algorithms described in the previous Section were tested using two sets of images from a vase and a head model, respectively (see figure 4.8). Both sets consisted of 36 images, with the turntable rotated by an angle of 10° between successive snapshots. The fact that the angle was fixed was not used either in the estimation of the epipolar geometry, nor in the reconstruction to be shown in Section 4.7. For the vase sequence, the symmetry transformation associated with the envelope of its profiles was assumed to be a harmonic homology \mathbf{W} , whereas for the head sequence the transformation was modelled as a skewed symmetry \mathbf{S} . The choice of the simpler model for the head sequence was motivated by the fact that the camera was nearly pointing towards the axis of the turntable, and therefore the skewed symmetry transformation could be used. Of course, there would have been no problem in adopting the more complex model. To obtain \mathbf{W} and \mathbf{S} , procedure 4.2 was implemented with

100 evenly spaced sample points along each envelope ($N = 100$). Initialisations were done by using bitangents. Less than 10 iterations of the Levenberg-Marquardt algorithm were necessary, with derivatives computed by finite differences. The final positions of the rotation axes can be seen in figure 4.9.

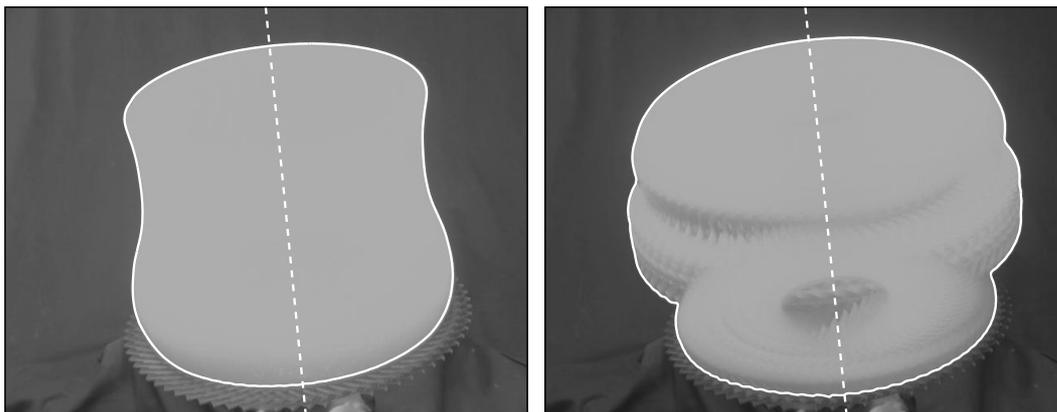


Figure 4.9. Overlap of the images of the vase (left) and of the head (right). The solid lines are the envelopes of the profiles and dashed lines are estimates of the images of the rotation axis in both sequences.

In the implementation of procedure 4.3, seventy pairs of images were selected by uniformly sampling the indexes of the images in each sequence, and the resultant estimate of the epipoles for the vase sequence is shown in figure 4.10, which also shows the horizon \mathbf{l}_h found by a robust fit to the epipoles. To get \mathbf{l}_h a minimisation of the median of the squares of the residuals was used, followed by removal of outliers and orthogonal least-squares regression using the remaining points (inliers). The epipolar geometry was then re-estimated with the epipoles constrained to lie on \mathbf{l}_h . Once the epipolar geometry was obtained, precomputed intrinsic parameters were used to convert the fundamental matrices into essential matrices, and these were then decomposed to provide the camera motion and orientation. The resulting camera configurations are presented in figure 4.11.

The object was rotated on a manual turntable with resolution of 0.01° , but the

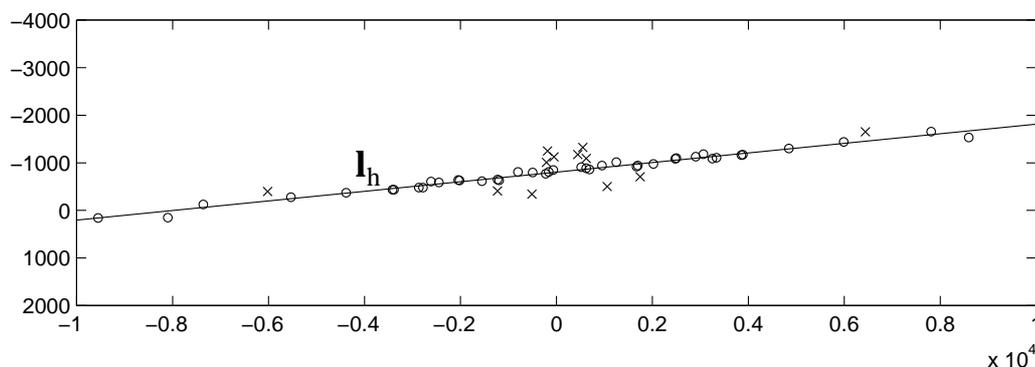


Figure 4.10. Epipoles estimated by procedure 4.3. The horizon was found by doing a robust fit to the cloud of epipoles. Inliers are shown as circles (\circ) and outliers as crosses (\times).

real precision achieved is highly dependent on the skill of the operator. The RMS errors in the estimated angles were 0.19° and 0.23° for the vase and head sequence respectively (see figure 4.12), demonstrating the accuracy of the estimation.

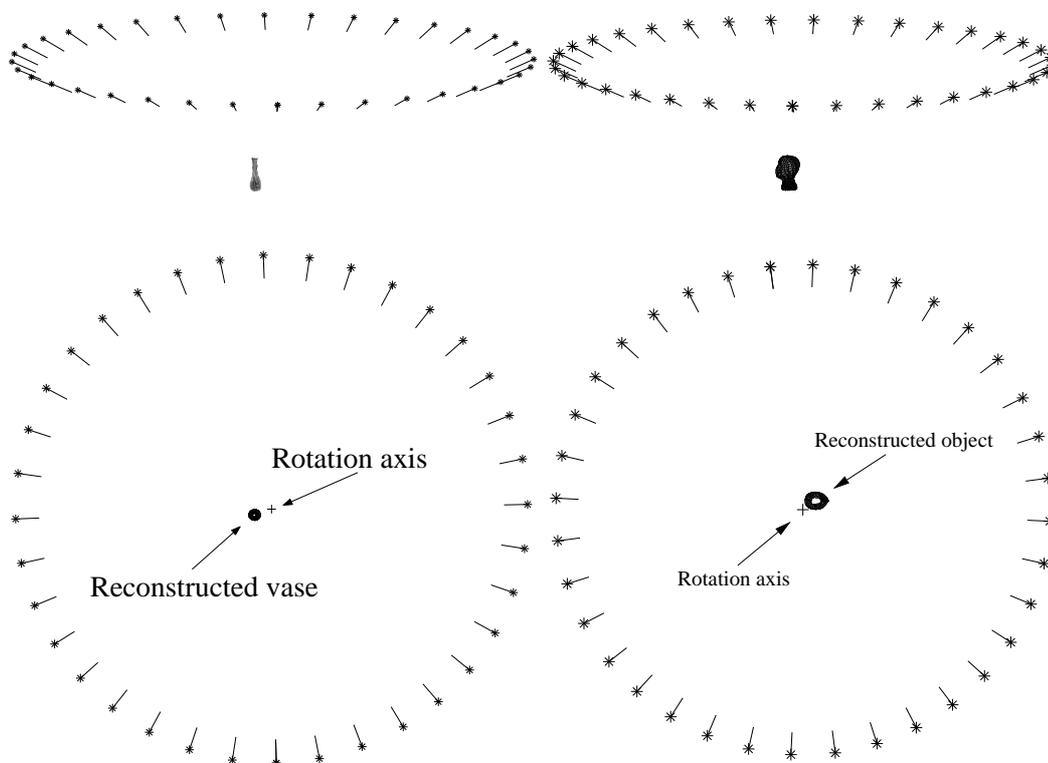


Figure 4.11. Camera configurations for the vase (left) and head (right) sequences.

It is interesting to compare this result with the ones shown in [50, pg. 166] for the “Head”, “Freiburg” and “Dinosaur” sequences, where the average number of point matches per image pair varies from 137 to 399, depending on the sequence. It should be stressed that only two epipolar tangents were used for each pair of images in the experiments presented in this paper, with comparable results.

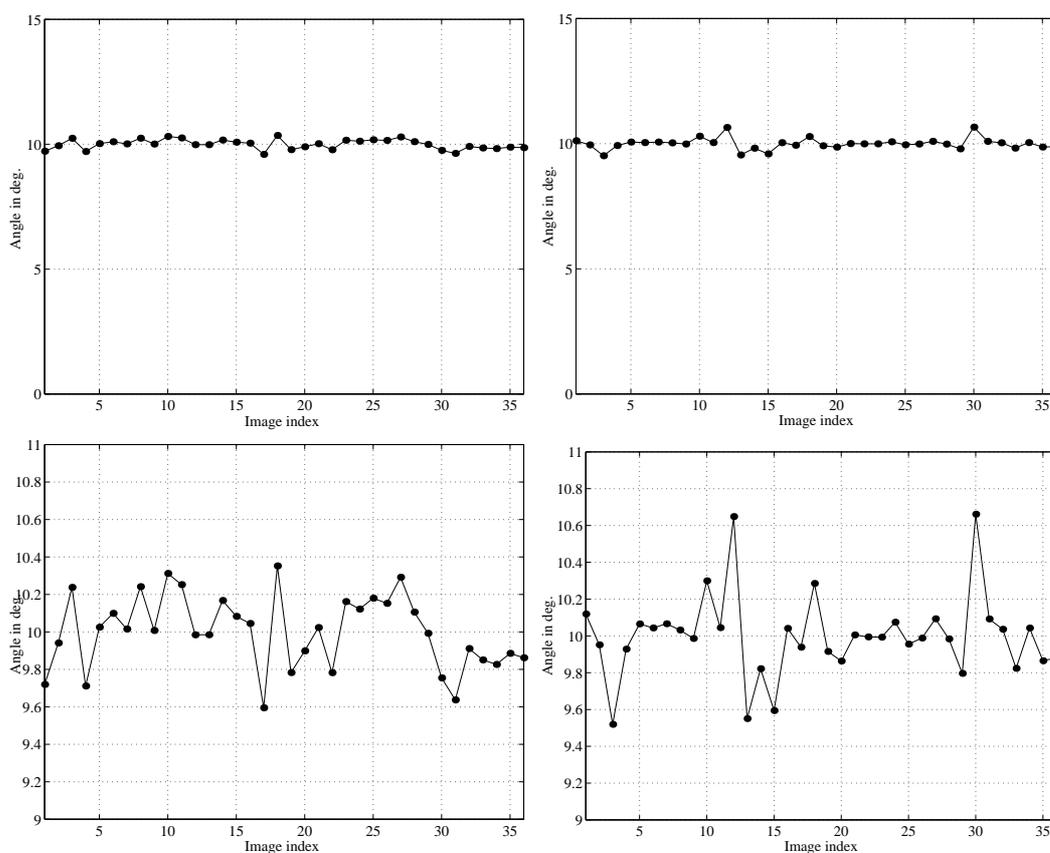


Figure 4.12. Estimated angles of rotation between successive views for the vase (left) and head (right) sequences, with RMS errors 0.19° and 0.23° , respectively.

4.7 Reconstruction from Image Profiles

The algorithm for motion estimation introduced here can be used even when point correspondences can be established. On the other hand, methods such as the ones in

Algorithm 4.2 Reconstruction from image profiles.

```

for  $i = 1$  to  $N - 1$  do
  sample  $M$  points  $\mathbf{u}_j$  along the profile in image  $i$ ;
  for  $j = 1$  to  $M$  do
    compute the epipolar line  $\mathbf{l}$  at image  $i + 1$  corresponding to the point  $\mathbf{u}_j$ ;
    find the intersection  $\mathbf{u}'_j$  of the line  $\mathbf{l}$  with the profile in image  $i + 1$ ;
    triangulate the points  $\mathbf{u}_j$  and  $\mathbf{u}'_j$ ;
  end for
end for

```

[147], [50] and [84] cannot deal with situations where profiles are the only available features in the scene, and it is therefore natural to use the motion recovered by the technique shown in this paper to perform reconstruction from profiles. To solve this problem under known motion, the main algorithms can be found in [146, 158, 27, 14, 161]. Results reported in [161] compare the last three, and although it slightly favours the one in [14], the simplicity of the method proposed in [161] justifies its choice for evaluating the accuracy of the motion estimated here. It should be clear, however, that once the camera motion is estimated, a number of techniques for reconstruction could be used, such as voxel-carving [145, 89] or level-set methods [43].

4.7.1 Description of the Method

The algorithm for reconstruction from profiles introduced in [161] is based on the assumption that, if the motion is small, the error in triangulating correspondences in images of successive contour generators, established via the epipolar parameterisation, will be negligible (see figure 4.13). This corresponds to a finite-difference approximation of the technique shown in [27]. A summary of the procedure is shown in algorithm 4.2.

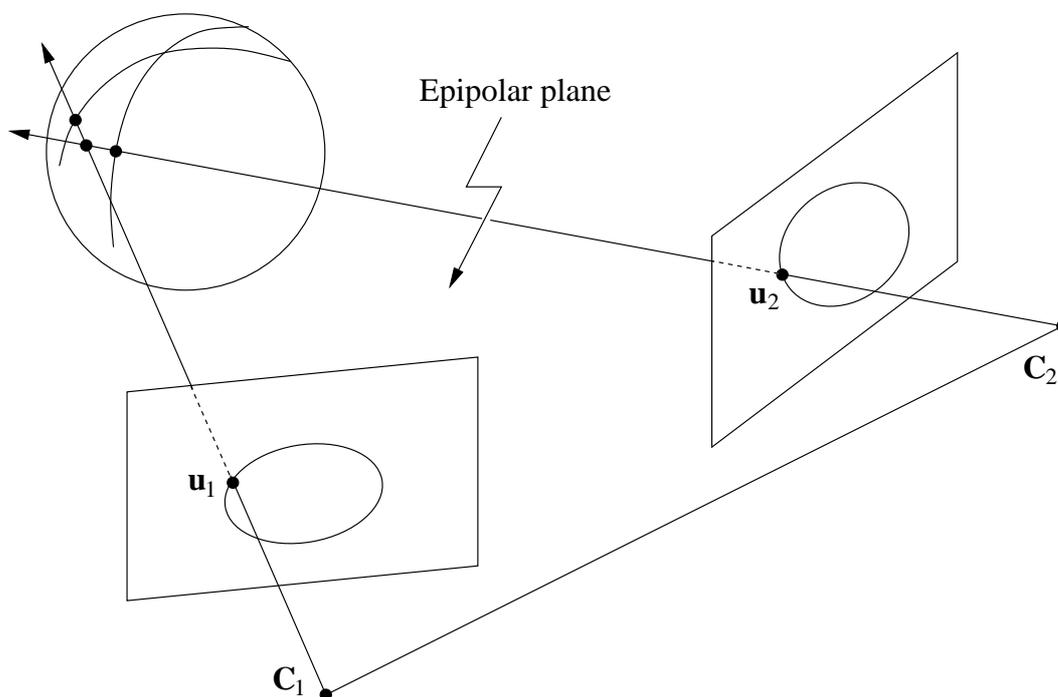


Figure 4.13. The correspondence between the points \mathbf{u}_1 and \mathbf{u}_2 is established via the epipolar parameterisation. The result of the triangulation of \mathbf{u}_1 and \mathbf{u}_2 is *not* a point on the surface, but if the motion is small, the error will be negligible.

4.7.2 Implementation and Experimental Results

B-splines were fitted to the left sides of the profiles in the sequences. From top to bottom, 18 points were sampled along the splines in the first image (see figure 4.14(a)), from which the corresponding epipolar lines in the second image were computed, and the corresponding points were then triangulated. The intersection of the epipolar lines with the profile at the second image is shown in figure 4.14(b). Since the corresponding points satisfy the epipolar constraint by construction, the triangulation will be exact, i.e., the rays associated with the points at the first image will exactly intersect the corresponding rays from the second image. As pointed out in [69], in this case the choice of triangulation method becomes irrelevant, and a simple least-squares solution was adopted. Details of the 3D reconstruction of the

object are shown in figure 4.15 and figure 4.16.

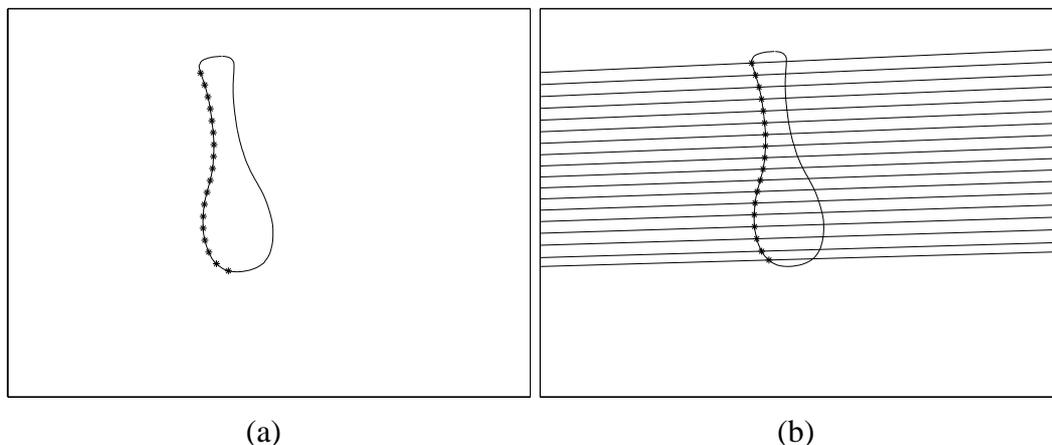


Figure 4.14. (a) Points sampled at the first image. (b) Corresponding epipolar lines at the second image. The triangulation is carried out between a point in the first image and the intersection of its corresponding epipolar line and the profile in the second image.

4.8 Summary and Conclusions

This chapter introduced a novel technique for motion estimation from image profiles. It does not make use of expensive search procedures, such as bundle adjustment, although it naturally integrates data from multiple images. The method is mathematically sound, practical and highly accurate. From the motion estimation to the model reconstruction, no point tracking is required and it does not depend on having point correspondences beforehand.

The convergence to local minima, a critical issue in most non-linear optimisation problems, is avoided by a divide-and-conquer approach that keeps the size of the problem manageable. Moreover, a search space with lower dimension results in fewer iterations before convergence. The quality of model reconstructed is remarkable, in particular if one considers that only the least possible amount of information has been used.

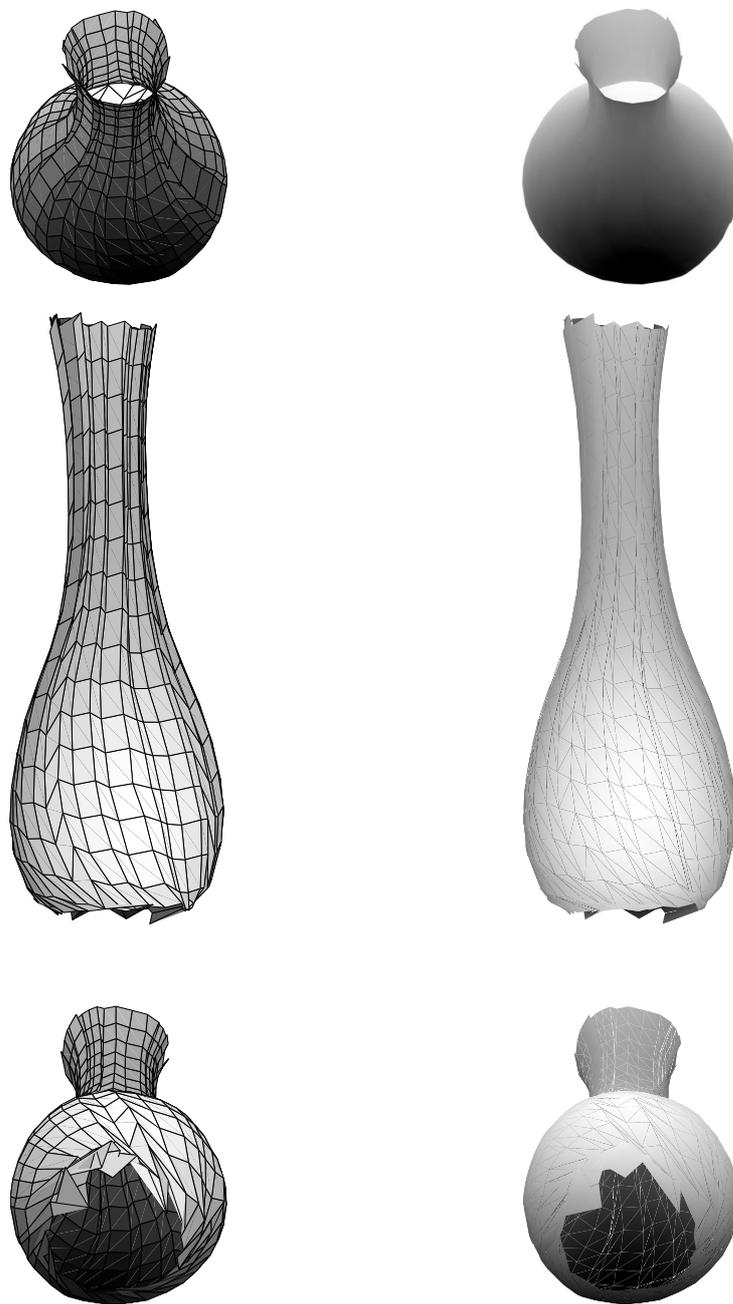


Figure 4.15. Details of the reconstruction of the vase. The left column shows the mesh and the right column shows the reconstruction after shading.

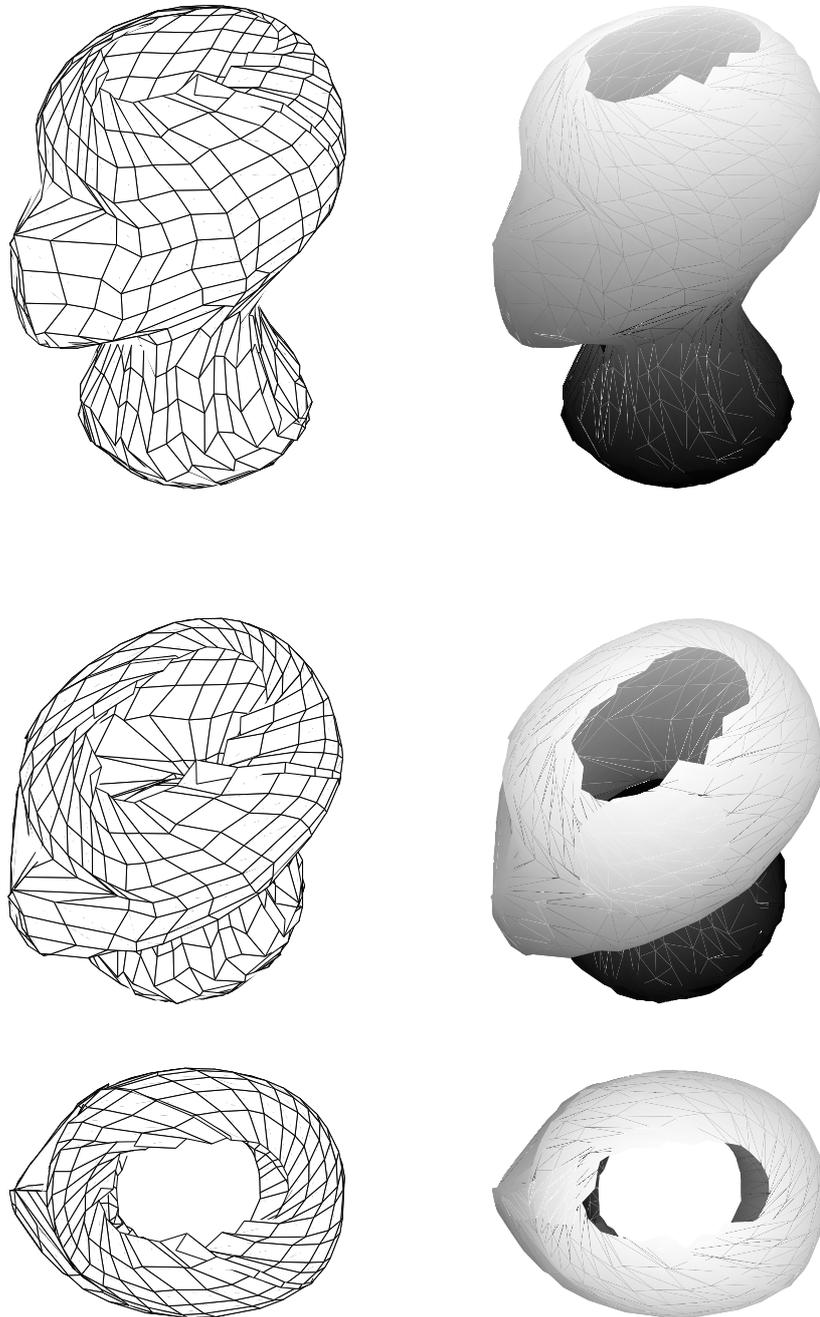


Figure 4.16. Details of the reconstruction of the head. The left column shows the mesh and the right column shows the reconstruction after shading.

Chapter 5

Camera Self-Calibration

5.1 Introduction

This chapter introduces a novel technique for self-calibration of cameras with possibly varying intrinsic parameters based on the Huang and Faugeras constraints, which state that the two nonzero singular values of an essential matrix must be equal. For known skew, aspect ratio and principal point, this condition is used to derive an approximate linear solution for the estimation of the focal lengths, which can be used to bootstrap a more accurate search for the parameters of the cameras. Under minimal assumptions, such as zero skew, the technique presented here can cope with variable intrinsic parameters, and it has a built-in detection of critical motions for self-calibration, which greatly improves the accuracy of the self-calibration.

The main contributions of the chapter are the development of a novel linear algorithm for estimating focal lengths of multiple cameras, the embedded detection of pairs of images for which the camera motion is critical for self-calibration [163, 141, 142, 80], and the introduction of a nonlinear refinement to the linear solution with simultaneous estimation of the remaining camera parameters that also

incorporates the critical motion detection. The input of the algorithm is a set of fundamental matrices, and it is not necessary to perform any kind of *projective factorisation* [143, 150, 72] or *projective bundle adjustment* [9, 143, 10] prior to the calibration. Experiments with synthetic and real data have shown that the technique is robust to noise and that it operates well under quasi-critical motions.

5.2 Previous Works

The problem of self-calibration has attracted the attention of researchers in the computer vision community for providing a powerful method for the recovery of 3D models from image sequences. Compared to the classical calibration problem [155, 49, 41], the algorithms for self-calibration make no or few assumptions about the particular structure of the scene being viewed. Instead, they attempt to calibrate the cameras by finding intrinsic parameters that are consistent with the underlying projective geometry of a sequence of images. This constraints of consistency can be expressed as the Kruppa equations [109, 101, 98], the Trivedi constraints [154], the Huang and Faugeras constraints [76, 61, 112], or formulated in terms of the *absolute quadric* [151, 122]. Interestingly, it has been shown that the Kruppa equations, the Trivedi constraints and the Huang and Faugeras constraints are equivalent [101]. However, as pointed out in [97], that does not mean that they will produce the same results when used in self-calibration algorithms. The Kruppa equations have the advantage of requiring the solution of systems of polynomial equations of lower order when compared with the Trivedi and the Huang and Faugeras constraints (but see section 5.4.1), which, however, do not make explicit use of epipoles, whose estimation is notoriously inaccurate.

The concept of self-calibration was introduced by Maybank, Faugeras and Lu-

ong [109, 44], who proposed an approach based on the Kruppa equations when up to three views of a scene are available, and established the relation between camera intrinsic parameters and the *absolute conic*. An algorithm for computing the focal lengths of two cameras given the corresponding fundamental matrix and knowledge of the remaining intrinsic parameters was provided by Hartley in [61] (an elegant closed-form solution for the same problem can be found in [13]). That paper also made the first use of the Huang and Faugeras constraints as a tool for self-calibration. In [62] the ideas in [109] were used in the development of a practical algorithm for self-calibration for more than three cameras. Together with [45], it also put forward the idea of finding an appropriate 3D homography that updates a projective reconstruction to a Euclidean one. This approach was further developed by Triggs in [151], where the absolute quadric was introduced, and by Pollefeys, Koch and Van Gool in [121, 122], where a practical method for self-calibration of multiple cameras with varying intrinsic parameters was developed.

The technique presented in this paper generalises the ones introduced in [61, 13]. It has a linear step, where only the focal lengths are computed, followed by a nonlinear optimisation that refines the estimate obtained in the linear stage and allows for the estimation of more intrinsic parameters, such as the principal point. Moreover, it naturally takes into account how close to a *critical motion for self-calibration* [163, 141, 142, 80] the relative motion between any two pair of cameras in the sequence is, and “weights” the information provided by that particular pair accordingly. Finally, the input of the algorithm is only a set of fundamental matrices, and therefore there is no need for projective bundle adjustment. This is an interesting advantage, for although it is easy to compute fundamental matrices consistent with a given set of projective camera matrices (and still use the technique proposed here),

the estimation of projective camera matrices given a set of fundamental matrices is a nontrivial problem, due to nonlinear relations the fundamental matrices must simultaneously satisfy [48].

5.3 Theoretical Background

The general form for the matrix \mathbf{K} of intrinsic parameters of a pinhole camera is

$$\mathbf{K} = \begin{bmatrix} \alpha & s & u \\ 0 & \varepsilon\alpha & v \\ 0 & 0 & 1 \end{bmatrix}. \quad (5.1)$$

It has been shown in chapter 3 that given the fundamental matrix \mathbf{F} related to a pair of images with intrinsic parameters given by \mathbf{K}_1 and \mathbf{K}_2 , the corresponding essential matrix will be given by $\mathbf{E} = \mathbf{K}_2^T \mathbf{F} \mathbf{K}_1$ [102]. The relative translation \mathbf{t} and rotation \mathbf{R} between the images can be found decomposing the essential matrix \mathbf{E} as

$$\mathbf{E} = [\mathbf{t}]_{\wedge} \mathbf{R}. \quad (5.2)$$

5.3.1 The Huang and Faugeras Constraints

Several works have pointed out the possibility of exploiting the Huang and Faugeras constraints for self-calibration [61, 112, 97]. This constraint states that the two nonzero singular values of an essential matrix \mathbf{E} must be equal [156, 76], i.e., $\mathbf{E} =$

\mathbf{UDV}^T , where

$$\mathbf{D} = \begin{bmatrix} \sigma & 0 & 0 \\ 0 & \sigma & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad (5.3)$$

and \mathbf{U} and \mathbf{V} are orthonormal matrices. This constraint is a necessary and sufficient condition for the decomposition of \mathbf{E} as shown in (5.2) to be possible.

5.4 Self-Calibration from the Essential Matrix

Consider a sequence of n images taken from n pinhole cameras (or from the same pinhole camera in n different positions), and let \mathbf{K}_i be the matrix of intrinsic parameters of camera i , where it is assumed zero skew and aspect ratio one (see section 2.3.1), i.e.,

$$\mathbf{K}_i = \begin{bmatrix} \alpha_i & 0 & u_i \\ 0 & \alpha_i & v_i \\ 0 & 0 & 1 \end{bmatrix}. \quad (5.4)$$

Therefore, the essential matrix related to images i and j is given by

$$\mathbf{E}_{ij} = \mathbf{K}_j^T \mathbf{F}_{ij} \mathbf{K}_i, \quad (5.5)$$

where \mathbf{F}_{ij} is the fundamental matrix corresponding to images i and j . Nevertheless, the Huang and Faugeras constraints will not be satisfied in (5.5) for arbitrary matrices \mathbf{K}_j and \mathbf{K}_i . Since any fundamental matrix has rank two, the rank of \mathbf{E}_{ij} will also be two for any \mathbf{K}_j and \mathbf{K}_i with full rank. However, as pointed out by

Longuet-Higgins and first mentioned in [76], the condition that two singular values of a matrix are equal yields two constraints on its elements, which have to be additionally imposed over \mathbf{K}_j and \mathbf{K}_i for (5.2) to be possible.

5.4.1 Linear Solution

This section presents the derivation of a novel linear algorithm for self-calibration under the assumptions of known principal point, aspect ratio and skew. It generalises for multiple cameras the results found in [61, 13], and further extends them to take into account critical and quasi-critical camera configurations, as will be shown in subsection 5.4.2. Under the assumption that so many intrinsic parameters are known, it is reasonable to argue that one could simply trust the (presumably) highly accurate specifications for the value of the focal length provided by the camera manufacturer. However, for a sequence of images acquired from a zooming camera, the focal length will be varying, despite any default value indicated by the manufacturer. In this situation the algorithm presented here can clearly play an important role.

A necessary condition for the validity of the Huang and Faugeras constraints derived by Longuet-Higgins [76] is that if

$$\mathbf{E}_{ij} = \begin{bmatrix} (\mathbf{b}_1^{ij})^T \\ (\mathbf{b}_2^{ij})^T \\ (\mathbf{b}_3^{ij})^T \end{bmatrix}^T, \quad (5.6)$$

then

$$\frac{\mathbf{b}_1^{ij}}{\mathbf{b}_2^{ijT} \mathbf{b}_3^{ij}} + \frac{\mathbf{b}_2^{ij}}{\mathbf{b}_1^{ijT} \mathbf{b}_3^{ij}} + \frac{\mathbf{b}_3^{ij}}{\mathbf{b}_1^{ijT} \mathbf{b}_2^{ij}} = \mathbf{0}. \quad (5.7)$$

This result can be directly verified by expanding (5.2) with $\mathbf{t} = [t_1 \ t_2 \ t_3]^T$ and $\mathbf{R} = [\mathbf{r}_1 \ \mathbf{r}_2 \ \mathbf{r}_3]^T$, from which one obtains

$$\mathbf{E}_{ij} = \begin{bmatrix} -t_3 \mathbf{r}_2^T + t_2 \mathbf{r}_3^T \\ +t_3 \mathbf{r}_1^T - t_1 \mathbf{r}_3^T \\ -t_2 \mathbf{r}_1^T + t_1 \mathbf{r}_2^T \end{bmatrix}^T, \quad (5.8)$$

and (5.7) follows by comparing (5.6) and (5.8). Moreover, (5.7) can be rewritten as

$$\begin{bmatrix} 1/(\mathbf{b}_2^{ijT} \mathbf{b}_3^{ij}) & 1/(\mathbf{b}_1^{ijT} \mathbf{b}_3^{ij}) & 1/(\mathbf{b}_1^{ijT} \mathbf{b}_2^{ij}) \end{bmatrix} \begin{bmatrix} (\mathbf{b}_1^{ij})^T \\ (\mathbf{b}_2^{ij})^T \\ (\mathbf{b}_3^{ij})^T \end{bmatrix} = \mathbf{0}, \quad (5.9)$$

and, therefore,

$$\begin{bmatrix} 1/(\mathbf{b}_2^{ijT} \mathbf{b}_3^{ij}) \\ 1/(\mathbf{b}_1^{ijT} \mathbf{b}_3^{ij}) \\ 1/(\mathbf{b}_1^{ijT} \mathbf{b}_2^{ij}) \end{bmatrix} \in \mathcal{LN}(\mathbf{E}_{ij}), \quad (5.10)$$

where $\mathcal{LN}(\mathbf{E}_{ij})$ is the left null space of \mathbf{E}_{ij} . However, from (5.5),

$$\mathcal{LN}(\mathbf{E}_{ij}) = \mathbf{K}_j^{-1} \mathbf{e}^{ji} \lambda^{ji}, \quad (5.11)$$

where \mathbf{e}^{ji} is the left epipole of \mathbf{F}_{ij} and $\lambda^{ji} \in \mathbb{R}$. Let $u_i = 0$ and $v_i = 0$ in (5.4) and

$$\mathbf{F}_{ij} = \begin{bmatrix} f_{11}^{ij} & f_{12}^{ij} & f_{13}^{ij} \\ f_{21}^{ij} & f_{22}^{ij} & f_{23}^{ij} \\ f_{31}^{ij} & f_{32}^{ij} & f_{33}^{ij} \end{bmatrix}. \quad (5.12)$$

Therefore, from (5.5),

$$\mathbf{E}_{ij} = \begin{bmatrix} \alpha_i \alpha_j f_{11}^{ij} & \alpha_i \alpha_j f_{12}^{ij} & \alpha_j f_{13}^{ij} \\ \alpha_i \alpha_j f_{21}^{ij} & \alpha_i \alpha_j f_{22}^{ij} & \alpha_j f_{23}^{ij} \\ \alpha_i f_{31}^{ij} & \alpha_i f_{32}^{ij} & f_{33}^{ij} \end{bmatrix}, \quad (5.13)$$

and

$$\begin{aligned} \mathbf{b}_2^{ijT} \mathbf{b}_3^{ij} &= \alpha_i^2 \alpha_j f_{21}^{ij} f_{31}^{ij} + \alpha_i^2 \alpha_j f_{22}^{ij} f_{32}^{ij} + \alpha_j f_{23}^{ij} f_{33}^{ij}, \\ \mathbf{b}_1^{ijT} \mathbf{b}_3^{ij} &= \alpha_i^2 \alpha_j f_{11}^{ij} f_{31}^{ij} + \alpha_i^2 \alpha_j f_{12}^{ij} f_{32}^{ij} + \alpha_j f_{13}^{ij} f_{33}^{ij}, \\ \mathbf{b}_1^{ijT} \mathbf{b}_2^{ij} &= \alpha_i^2 \alpha_j^2 f_{11}^{ij} f_{21}^{ij} + \alpha_i^2 \alpha_j^2 f_{12}^{ij} f_{22}^{ij} + \alpha_j^2 f_{13}^{ij} f_{23}^{ij}. \end{aligned} \quad (5.14)$$

Thus, substituting (5.10), (5.11) and (5.13) in (5.14), with $\mathbf{e}^{ji} = [e_1^{ji} \ e_2^{ji} \ e_3^{ji}]^T$, one obtains

$$\alpha_j / (e_1^{ji} \lambda^{ji}) = \alpha_i^2 \alpha_j f_{21}^{ij} f_{31}^{ij} + \alpha_i^2 \alpha_j f_{22}^{ij} f_{32}^{ij} + \alpha_j f_{23}^{ij} f_{33}^{ij}, \quad (5.15)$$

$$\alpha_j / (e_2^{ji} \lambda^{ji}) = \alpha_i^2 \alpha_j f_{11}^{ij} f_{31}^{ij} + \alpha_i^2 \alpha_j f_{12}^{ij} f_{32}^{ij} + \alpha_j f_{13}^{ij} f_{33}^{ij}, \quad (5.16)$$

$$1 / (e_3^{ji} \lambda^{ji}) = \alpha_i^2 \alpha_j^2 f_{11}^{ij} f_{21}^{ij} + \alpha_i^2 \alpha_j^2 f_{12}^{ij} f_{22}^{ij} + \alpha_j^2 f_{13}^{ij} f_{23}^{ij}. \quad (5.17)$$

The focal length α_i can be obtained from (5.15) and (5.16) by solving the system of linear equations

$$\begin{bmatrix} (f_{11}^{ij} f_{31}^{ij} + f_{12}^{ij} f_{32}^{ij}) & -1/e_2^{ji} \\ (f_{21}^{ij} f_{31}^{ij} + f_{22}^{ij} f_{32}^{ij}) & -1/e_1^{ji} \end{bmatrix} \begin{bmatrix} \alpha_i^2 \\ 1/\lambda^{ji} \end{bmatrix} = \begin{bmatrix} -f_{13}^{ij} f_{33}^{ij} \\ -f_{23}^{ij} f_{33}^{ij} \end{bmatrix} \Leftrightarrow \mathbf{A}_{ij} \mathbf{x}_{ij} = \mathbf{b}_{ij}. \quad (5.18)$$

The equations in (5.18) provide a linear solution for the estimation of the focal length. Interestingly, the computation of α_i according to (5.18) gives the same

expression found in [13] when \mathbf{p} and \mathbf{p}' (corresponding to the principal points in that author's notation) are made equal to $[0 \ 0 \ 1]^T$.

5.4.2 Detection of Critical Motions

It has been pointed out in several works [163, 141, 142] that, for some particular camera motions, self-calibration of all parameters may not be possible. The addition of further constraints — such as zero skew, square pixels or knowledge of the principal point — may resolve the ambiguities that arise under such conditions. However, some particular configurations remain ambiguous even when these constraints are imposed. In particular, it has been shown in [166] that, when the camera motion is restricted to an arbitrary translation followed by a rotation about a line perpendicular to the image plane, no combination of the constraints mentioned above provides enough information for the resolution of the ambiguity. In fact, it has been shown in [80] that, for a pair of cameras with varying focal length this ambiguity remains for the larger class of motions where all pairs of optical axes intersect each other. Observe that the motion described in [166] is a particular case of this one, where the optical axes intersect at the same point at infinity.

Let \mathbf{P}_1 and \mathbf{P}_2 , where

$$\mathbf{P}_1 = \mathbf{K}_1[\mathbb{I} \ \mathbf{0}] \text{ and } \mathbf{P}_2 = \mathbf{K}_2[\mathbf{R} \ \mathbf{t}], \quad (5.19)$$

be the camera matrices of two projective cameras, and assume the aspect ratio, skew and principal points, denoted as \mathbf{u}_1 and \mathbf{u}_2 , are known. Without loss of generality,

we can make

$$\mathbf{K}_i = \begin{bmatrix} \alpha_i & 0 & 0 \\ 0 & \alpha_i & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (5.20)$$

Moreover, assume that the optical axes of \mathbf{P}_1 and \mathbf{P}_2 intersect. Therefore, the optical axes define an epipolar plane whose corresponding epipolar lines pass through \mathbf{u}_1 and \mathbf{u}_2 (see figure 5.1). This means that, if \mathbf{F} is the fundamental matrix relating the cameras, with element ij denoted by f_{ji} , the principal points must satisfy the epipolar constraint, i.e., $\mathbf{u}_2^T \mathbf{F} \mathbf{u}_1 = 0$. However, it is assumed that $\mathbf{u}_1 = \mathbf{u}_2 = [0 \ 0 \ 1]^T$, and, hence, $f_{33} = 0$. However, from [102, Section 2.3] and assuming $\mathbf{t} = [t_1 \ t_2 \ t_3]^T$, the fundamental matrix relating the cameras in (5.19) is given by

$$\mathbf{F} = [\mathbf{K}_2 \mathbf{t}]_{\wedge} \mathbf{K}_2 \mathbf{R} \mathbf{K}_1^{-1}. \quad (5.21)$$

If the element ij of \mathbf{R} is denoted by r_{ij} , the expansion of f_{33} in (5.21) gives $t_1 r_{23} - t_2 r_{13} = 0$, or

$$t_1 = w r_{13} \text{ and } t_2 = w r_{23}, \quad (5.22)$$

for a given $w \in \mathbb{R}$. Substituting (5.22) in (5.21), we have

$$\mathbf{F} = \begin{bmatrix} wr_{23}r_{31} - r_{21}t_3 & wr_{23}r_{32} - r_{22}t_3 & \alpha_1 r_{23}(wr_{33} - t_3) \\ r_{11}t_3 - wr_{13}r_{31} & r_{12}t_3 - wr_{13}r_{32} & \alpha_1 r_{13}(t_3 - wr_{33}) \\ w\alpha_2(r_{13}r_{21} - r_{11}r_{23}) & w\alpha_2(r_{13}r_{22} - r_{12}r_{23}) & 0 \end{bmatrix}, \quad (5.23)$$

$$\mathbf{e}' = \begin{bmatrix} w\alpha_2 r_{13} \\ w\alpha_2 r_{23} \\ t_3 \end{bmatrix}. \quad (5.24)$$

Substituting (5.23) and (5.24) in (5.18) we finally obtain

$$\mathbf{A} = - \begin{bmatrix} w\alpha_2 r_{13} t_3 & 1/(w\alpha_2 r_{23}) \\ w\alpha_2 r_{23} t_3 & 1/(w\alpha_2 r_{13}) \end{bmatrix} \text{ and } \mathbf{b} = \mathbf{0}. \quad (5.25)$$

Since $\det(\mathbf{A}) = 0$, there will be a family of solutions for $[\alpha_1^2 \ 1/\lambda]^T$ in (5.18), given by $\mu \mathbf{n}_A$, where $\mu \in \mathbb{R}$ and $\mathbf{n}_A = [n_1 \ n_2]^T$ is a unitary vector in the right null space of \mathbf{A} . By using (5.17) and (5.23), it is possible to show that the corresponding solution for α_2^2 will be $\alpha_2^2 = \mu n_2 / t_3 (\mu n_1 (f_{11} f_{21} + f_{12} f_{22}) + f_{13} f_{23})$.

If more cameras are added, but their relative displacement is such that any two optical axes intersect, all \mathbf{A}_{ij} in (5.18) will be rank deficient. Thus the condition number of \mathbf{A}_{ij} can be used as a heuristic measure of how close to critical the motion is. Let κ_{ij} be the inverse of the condition number of \mathbf{A}_{ij} . Therefore, when multiple images are available, the linear estimation of the focal length α_i of image i can be

As an example, assume that three images indexed by $i = 1, 2, 3$ are available. The three corresponding fundamental matrices are \mathbf{F}_{12} , \mathbf{F}_{13} and \mathbf{F}_{23} . The system of equations for the computation of the focal length of the first camera, α_1 , will then be given by

$$\begin{bmatrix} \kappa_{12}(f_{11}^{12} f_{31}^{12} + f_{12}^{12} f_{32}^{12}) & -\frac{\kappa_{12}}{e_2^{21}} & 0 \\ \kappa_{12}(f_{21}^{12} f_{31}^{12} + f_{22}^{12} f_{32}^{12}) & -\frac{\kappa_{12}}{e_1^{21}} & 0 \\ \kappa_{13}(f_{11}^{12} f_{31}^{13} + f_{12}^{12} f_{32}^{13}) & 0 & -\frac{\kappa_{13}}{e_2^{31}} \\ \kappa_{13}(f_{21}^{12} f_{31}^{13} + f_{22}^{12} f_{32}^{13}) & 0 & -\frac{\kappa_{13}}{e_1^{31}} \end{bmatrix} \begin{bmatrix} \alpha_1^2 \\ 1/\lambda^{21} \\ 1/\lambda^{31} \end{bmatrix} = \begin{bmatrix} -\kappa_{12} f_{13}^{12} f_{33}^{12} \\ -\kappa_{12} f_{23}^{12} f_{33}^{12} \\ -\kappa_{13} f_{13}^{13} f_{33}^{13} \\ -\kappa_{13} f_{23}^{13} f_{33}^{13} \end{bmatrix}. \quad (5.27)$$

A simple experiment was designed to validate the usefulness of κ in (5.26) as a measure of how close the motion of a given pair of cameras is to a critical configuration. The vergence angle v (i.e., the angle between the optical axes) of a stereo pair with intersecting optical axis was set to 1° , 5° , 10° and 15° . For each value of v the elevation angle τ of the second camera (i.e., the angle between its optical axis and the plane containing the two camera centers and the optical axis of the first camera) is varied in the range $[0^\circ, 30^\circ]$. The fundamental matrix relating the cameras for every combination of v and τ was then computed, followed by the matrix \mathbf{A} in (5.18). The inverse of the condition number of \mathbf{A} is shown in figure 5.2(a). For comparison, another measure of the criticality of the motion was computed, given by the geometric error of the respective principal points when matched through the epipolar constraint (see figure 3.2). The values of this alternative measure, denoted m , are shown in figure 5.2(b).

It can be seen from figure 5.2 that κ is approximately linear with the value of τ , the elevation angle, which is the main parameter controlling how close to critical the camera motion is. It also grows with v , the vergence angle, giving more weight

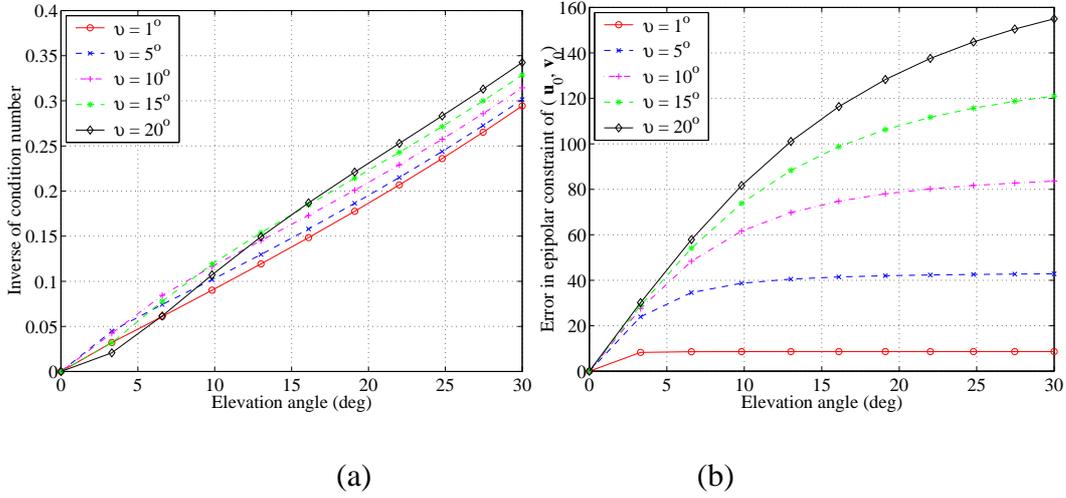


Figure 5.2. Measures of criticality of the camera motion. (a) Inverse of condition number of matrix \mathbf{A} in 5.18 (κ) and (b) geometric error in pixels of the reprojection of the principal point. Both (a) and (b) are plotted versus the elevation angle τ of the second camera, for several values of the vergence angle ν .

to larger baselines than to small ones. The alternative measure m , however, is not so well-behaved. Although it consistently increases with τ , it quickly saturates as κ grows. The use of m instead of κ in (5.26) would result in giving nearly the same weight to equations derived from cameras with the same ν , which is obviously undesired.

5.4.3 Nonlinear Solution

Once an initial approximation for the focal lengths of the cameras has been found by the linear algorithm, a full nonlinear optimisation can be run. If

$$\text{svd}(\mathbf{E}_{ij}) = \mathbf{U}_{ij} \begin{bmatrix} \sigma_1^{ij} & 0 & 0 \\ 0 & \sigma_2^{ij} & 0 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{V}_{ij}^T, \quad (5.28)$$

then, for \mathbf{E}_{ij} to be a valid essential matrix,

$$1 - \sigma_2^{ij}/\sigma_1^{ij} = 0 \quad (5.29)$$

must hold. Therefore, it is possible to search for the intrinsic parameters by minimising the cost function $\Lambda(\mathbf{K}_i, i = 1, \dots, n)$ given by

$$\Lambda(\mathbf{K}_i, i = 1, \dots, n) = \sum_{i=1}^n \sum_{j=i+1}^n \kappa_{ij} \left(1 - \frac{\sigma_2^{ij}}{\sigma_1^{ij}} \right). \quad (5.30)$$

The parameters of this cost function do not need to be only the focal lengths, but may also include the principal points or aspect ratios, as long as enough images are available. The expression in (5.30) is similar to the one in [112], but now the equations for each pair of camera are weighted according to the suitability of their relative motion regarding its closeness to a critical configuration for self-calibration.

Since the Huang and Faugeras constraints provide two equations on the intrinsic parameters per fundamental matrix, a naive computation would suggest that, because between n images there are $n(n-1)/2$ fundamental matrices, six images would allow for the calibration of *all* intrinsic parameters even if these were varying, which is clearly impossible. The contradiction arises from the fact that the $n(n-1)/2$ fundamental matrices are not independent, as discussed in section 5.2. The minimum number of images for the self-calibration of a camera where p_k parameters are known and p_f parameters are fixed, with the remaining $5 - p_k - p_f$ unknown and varying, was pointed out in [122] as n such that

$$n \times p_k + (n - 1) \times p_f \geq 8. \quad (5.31)$$

Typically, the skew is zero and the aspect ratio is one, and therefore $p_k = 2$. If the principal point and the focal length are unknown and varying, $p_f = 0$. Substituting the values of p_k and p_f in (5.31), one obtains $n \geq 4$, and thus self-calibration is possible with a minimum of four cameras. Under many practical situations [13] the principal point cannot be accurately estimated, and it can be safely assumed to be at the centre of the image. In this case, $p_k = 4$, and therefore the minimum number of cameras for self-calibration of the varying focal length is two.

It is important to observe that although the $n(n - 1)/2$ fundamental matrices associated with n cameras are not independent, it is beneficial to use the information provided by as many fundamental matrices as possible even when this information is redundant, since this procedure spreads the error more evenly across the cameras and the redundancy improves the robustness of the algorithm.

Alternative Expressions for the Huang and Faugeras Constraints There are several algebraic representations of the Huang and Faugeras constraints that do not make explicit use of the singular value decomposition of the essential matrix, in contrast to (5.29). A few examples are

$$\text{trace}^2(\mathbf{E}\mathbf{E}^T) - 2\text{trace}((\mathbf{E}\mathbf{E}^T)^2) = 0, \quad (5.32)$$

shown in [41], and, using the notation of (5.6),

$$\|\mathbf{b}_1^{ij} \times \mathbf{b}_2^{ij}\|^2 + \|\mathbf{b}_2^{ij} \times \mathbf{b}_3^{ij}\|^2 + \|\mathbf{b}_1^{ij} \times \mathbf{b}_3^{ij}\|^2 = \frac{1}{4}(\|\mathbf{b}_1^{ij}\|^2 + \|\mathbf{b}_2^{ij}\|^2 + \|\mathbf{b}_3^{ij}\|^2), \quad (5.33)$$

presented in [76]. In both cases the rank condition has yet to be imposed. There is a computational advantage in using the expressions given in (5.32) or (5.33) over

the one in (5.29), since the former ones use about 1/3 and 1/4, respectively, of the number of floating point operations necessary for the computation of (5.29) [58]. However, the computation of the derivatives of the cost function is much simpler if (5.29) is used. As will be discussed in the next session, this advantage justifies the choice of (5.29) as the algebraic expression of the Huang and Faugeras constraint to be used in the self-calibration algorithm.

5.5 Description of the Algorithm

Algorithm 5.1 Linear least-squares estimation of the focal lengths.

Assume that the principal points are at the centre of the images and transform the coordinate systems so that the principal points become $(0, 0)$;

for $i = 1$ to # images **do**

for $j = 1$ to # images **do**

if F_{ij} is defined **then**

 fill $\mathcal{A}_{2k-1,i}$, $\mathcal{A}_{2k,i}$, $\mathcal{A}_{2k-1,l}$, $\mathcal{A}_{2k,l}$, \mathcal{B}_{2k-1} and \mathcal{B}_{2k} according to (5.26);

end if

end for

 solve $\mathcal{AX} = \mathcal{B}$; $\alpha_i = \sqrt{\bar{\alpha}_i}$

end for

The ideas developed in the previous sections can be encapsulated into a practical and flexible algorithm for self-calibration. Consider a sequence of n images acquired from a camera with, possibly, varying intrinsic parameters. Initially, the fundamental matrices relating as many pairs of images as possible are computed. These fundamental matrices are then used to compute an initial guess for the focal lengths according to the linear algorithm described in (5.26) and summarised in algorithm 5.1. A more accurate solution can then be found by minimising the cost function shown in (5.30), as described in algorithm 5.2. Several configurations are possible, combining known, fixed and variable principal points with fixed and

variable focal lengths. To minimise the cost function in (5.30) any gradient based algorithm can be used, such as BFGS [15, 51, 57, 134] or Levenberg-Marquardt [93, 106] (for details of the implementation of the methods see [99, 126]). Experimental analysis has shown that the accuracy and convergence speed of the algorithm are greatly increased if the derivatives of the cost function are computed analytically instead of being estimated by finite differences. Closed-form solutions for the derivatives of (5.29) are provided in appendix B.

Algorithm 5.2 Nonlinear estimation of the intrinsic parameters.

Compute fundamental matrices between image pairs;
 initialise focal lengths using algorithm 5.1;
 initialise principal points at the centre of the images;
while not converged **do**
 compute cost function Λ in (5.30);
 compute $\frac{d\Lambda}{d\text{vec } \mathbf{K}_i}$ and $\frac{d\Lambda}{d\text{vec } \mathbf{K}_j}$ analytically using (B.5) and (B.6);
 update \mathbf{K}_i and \mathbf{K}_j to minimise Λ ;
end while

5.6 Experimental Results

To evaluate the robustness and accuracy of the algorithms proposed, experiments with both synthetic and real data were carried out. In all the experiments the skew was assumed to be zero and the aspect ratio was assumed to be one. In different experiments, the principal point was considered to be known, unknown but fixed, and varying.

5.6.1 Synthetic Data

To investigate the algorithm's robustness to noise, 50 points were randomly scattered in the interior of a cube centred at $(0, 0, 0)$ with edge of length one. Ten

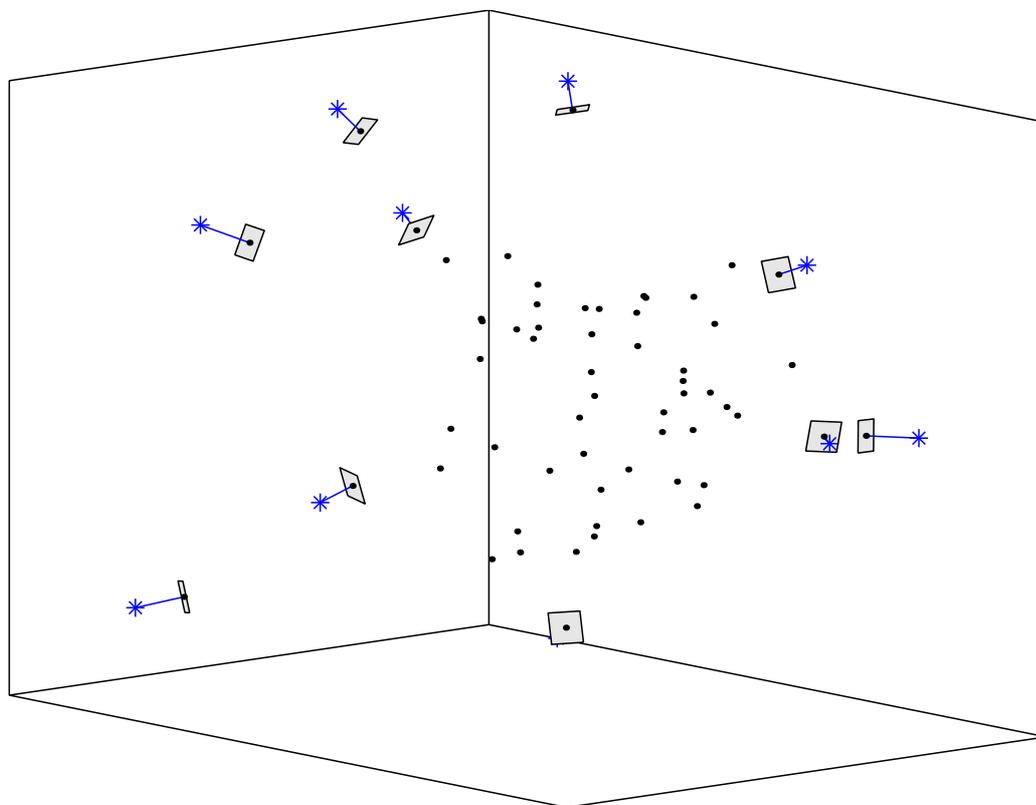


Figure 5.3. Cameras and points used in the self-calibration experiment with synthetic data.

cameras with principal points $(320, 240)$ and focal lengths from 600 to 2000 were placed around the cube, configured as shown in figure 5.3. Gaussian noise with different standard deviations was added to the image points, and all fundamental matrices between pairs of images on each group of 4 adjacent images were computed by minimising the reprojection error [102, 164]. The cameras were then calibrated using the linear algorithm, followed by the nonlinear algorithm assuming known, fixed and variable focal lengths. Results are shown in figures 5.4–5.7. The algorithm is highly accurate, estimating the focal length with about 1% of the true value when the image noise is one pixel. It can also be seen from figures 5.4–5.7

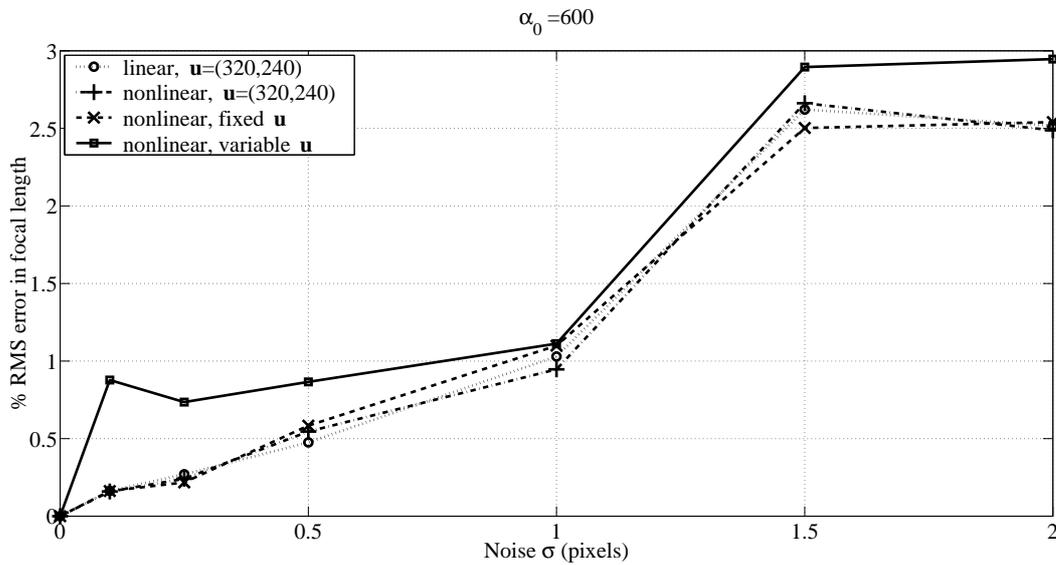


Figure 5.4. Percentage RMS errors for the focal length $\alpha_1 = 600$ for 10 experiments with synthetic data and different standard deviations σ of the image noise.

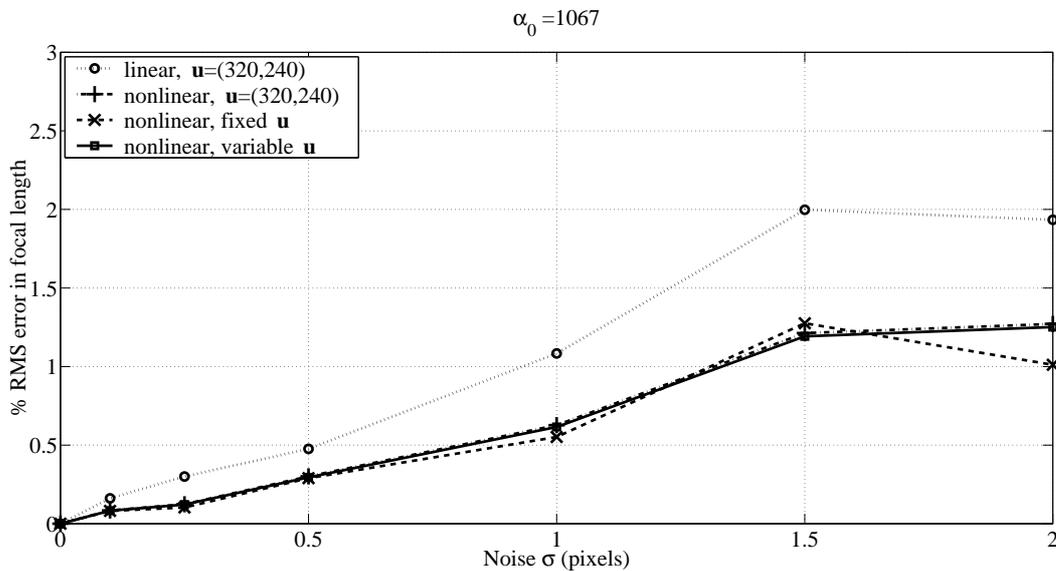


Figure 5.5. Percentage RMS errors for the focal length $\alpha_4 = 1067$ for 10 experiments with synthetic data and different standard deviations σ of the image noise.

that the error in the estimation degrades gracefully as the standard deviation of the image noise increases. Moreover, the relative accuracy of the nonlinear estimation

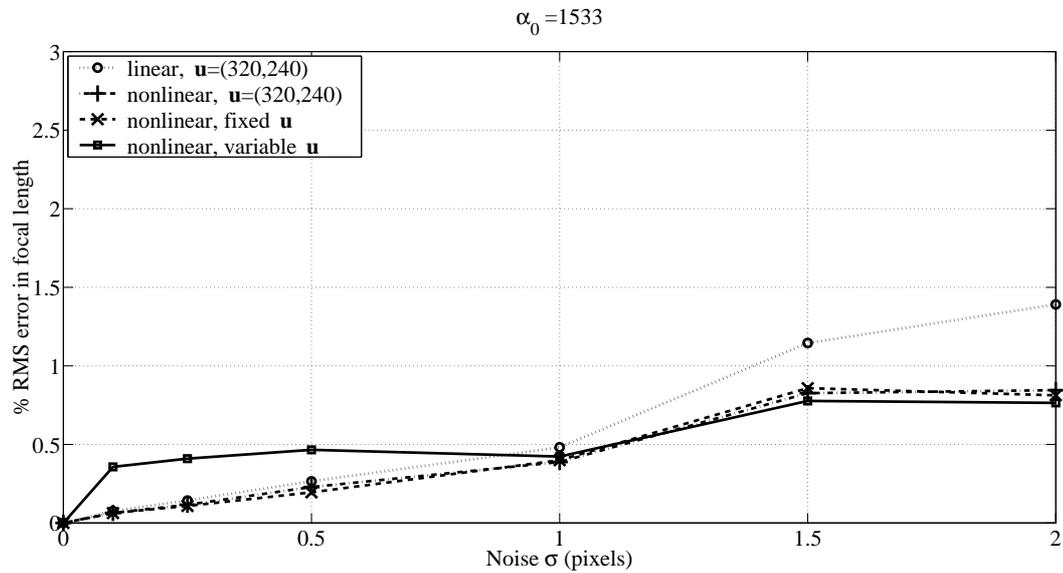


Figure 5.6. Percentage RMS errors for the focal length $\alpha_7 = 1533$ for 10 experiments with synthetic data and different standard deviations σ of the image noise.

is not significantly affected by the magnitude of the true focal length, since from figure 5.4 to figure 5.7 there was more than a threefold increase in this value with no corresponding change in the percentage RMS error. This effect is clearer in figure 5.8, which presents the tracking of the focal lengths throughout the images. The results shown are the average of 10 experiments.

5.6.2 Real Data

Six images of a calibration grid, shown in figure 5.9, were acquired with a digital camera with a resolution of 640×480 pixels. Three of the images were taken with a zoom factor of two, if the indications of the manufacturer can be trusted. The camera was then calibrated using the metric information of the calibration grid and a bundle adjustment algorithm [41, 153, 70]. Self-calibration was carried out using the linear algorithm, and the nonlinear algorithm assuming known principal point,

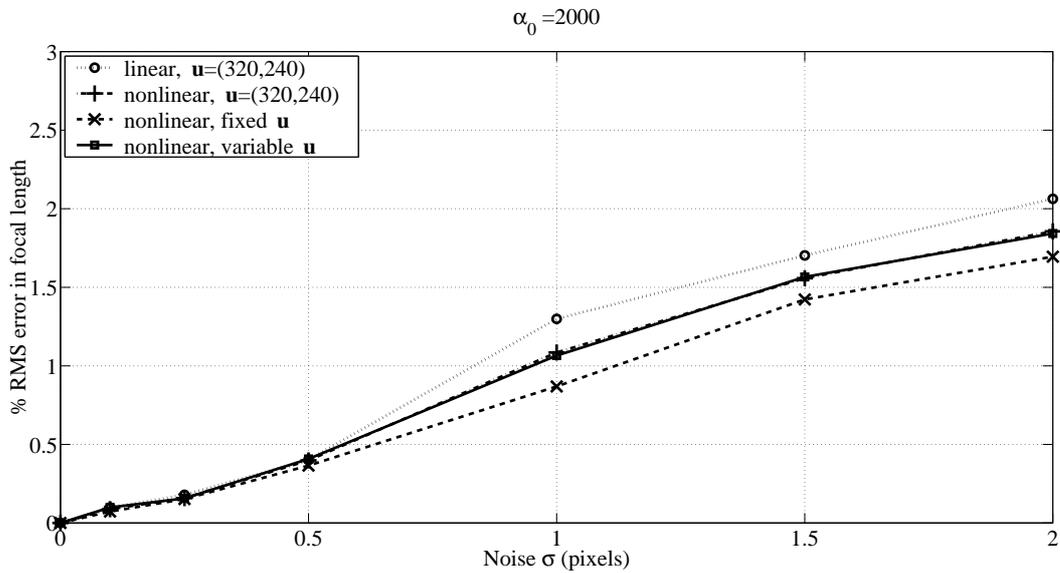


Figure 5.7. Percentage RMS errors for the focal length $\alpha_{10} = 2000$ for 10 experiments with synthetic data and different standard deviations σ of the image noise.

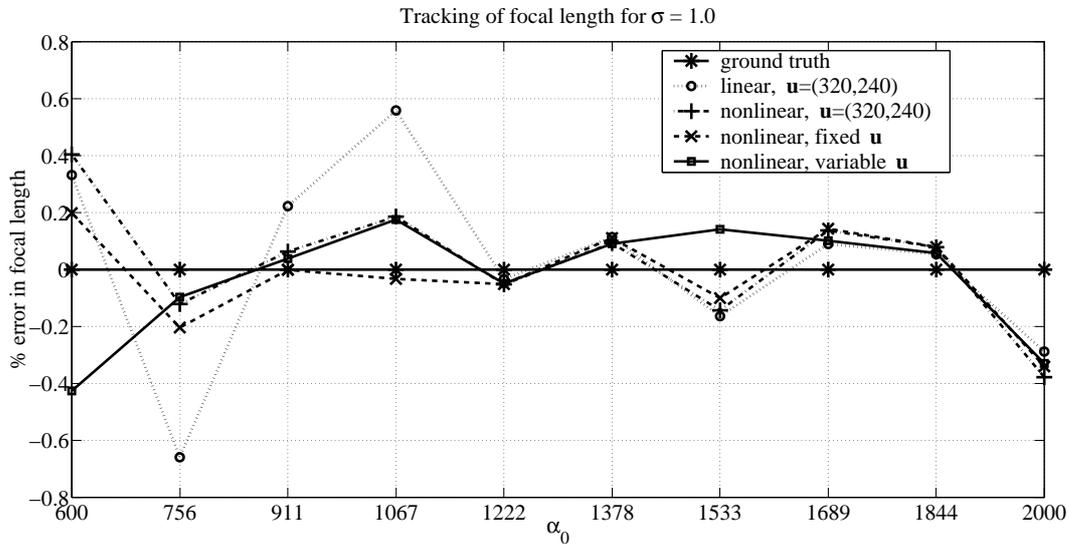


Figure 5.8. Tracking of the focal length. The results shown are the percentage average error of 10 experiments, with σ equal to 1 pixel.

unknown but fixed principal point, and variable principal point. Observe that the motion is quasi-critical, since all the cameras are roughly pointing at the same point

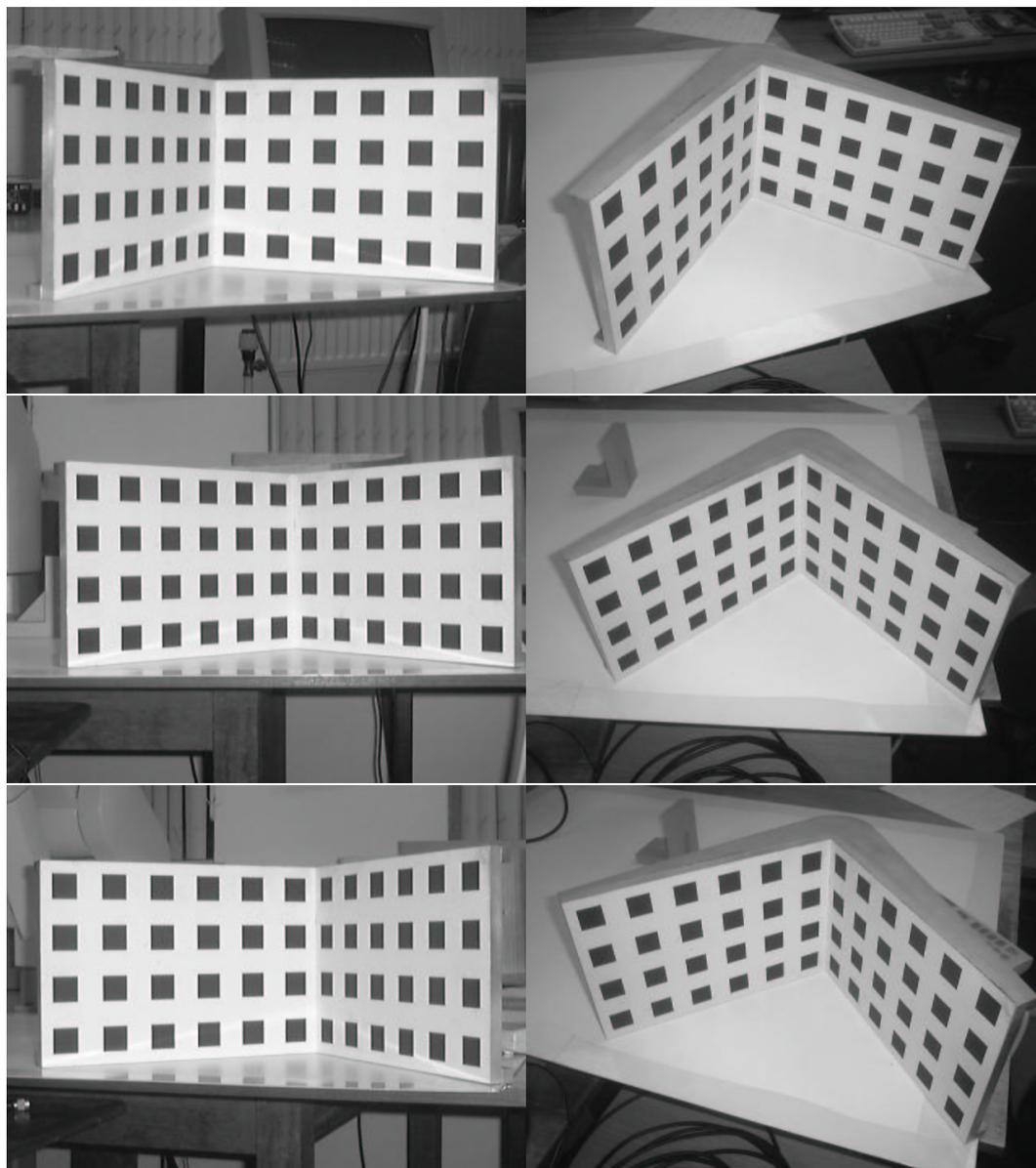


Figure 5.9. Images of the calibration grid used for experiments with real data

in space at the centre of the calibration grid.

Figure 5.10 shows the values of the focal lengths in different images. The linear algorithm is inaccurate, but its estimation of the focal length is good enough to be used as an initial guess for the nonlinear algorithm, which converges to a good solution whatever assumptions are made about the principal point. Figure 5.11

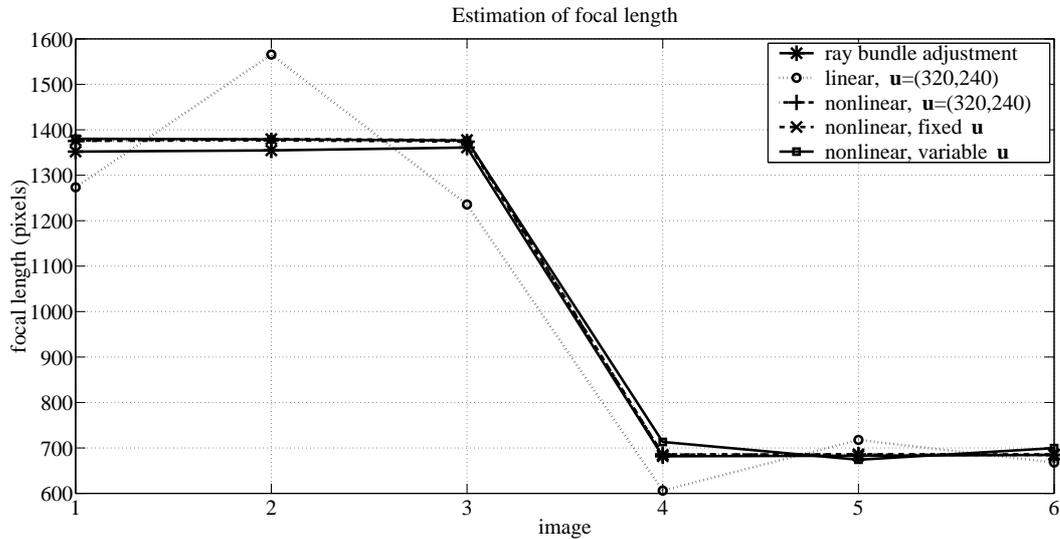


Figure 5.10. Estimated values of focal length from different algorithms.

shows the percentage error on the estimate of the focal lengths for the nonlinear algorithm. The error for each image is normalised by the focal length provided by the bundle adjustment algorithm. The three alternative assumptions for the principal point produce approximately the same results for the larger focal lengths in images one, two and three. Nevertheless, allowing for the principal point to vary clearly produces a worse result for the estimate of the smaller focal lengths in images four, five and six. It has been pointed out in [122] that for noise levels of the order of one pixel and above the simpler models should be preferred. To explain why the choice of the model is less critical when the focal length is larger, we turn our attention to figure 5.12.

Clearly, the principal point is not accurately estimated when it is allowed to vary, if we are to trust the result of the bundle adjustment algorithm. However, for larger values of the focal length, a precise estimation of the principal point is less significant, as shown in [13]. Therefore, the smaller values of the focal length are the ones that are most strongly affected by inaccuracies in the position of the

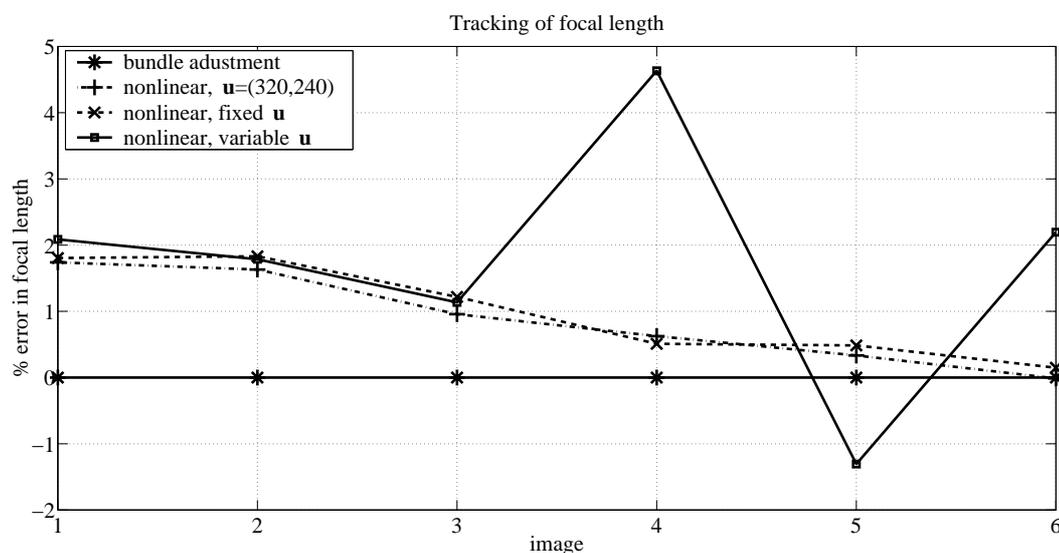


Figure 5.11. Percentage error in the focal lengths for different algorithms. If the principal point is allowed to vary the estimation of the focal length becomes less accurate.

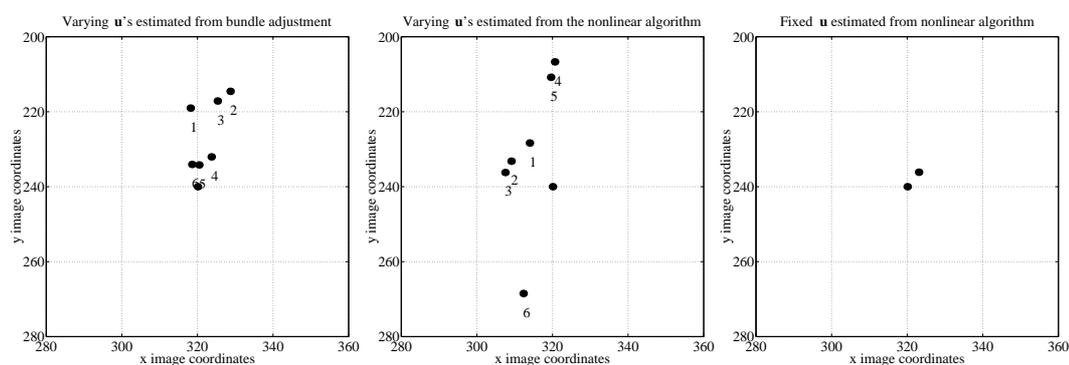


Figure 5.12. Principal points computed from different algorithms. It is clear that the principal point is not accurately estimated when it is allowed to vary. However, the effect of errors in the position of the principal point may be neglected when the focal length is larger.

principal point, producing the effect shown in figure 5.11.

A second experiment with real data was carried out, this time with a sequence of five outdoor images, of which four are shown in figure 5.13. Corners were detected by using a Harris corner detector [60, 59], and matched through correlation techniques [94, 82]. The algorithm described in section 3.1.4 was applied to estimate

the fundamental matrices relating the images, and the self-calibration algorithm introduced here was used to calibrate the cameras. The principal point was assumed to be fixed at the centre of the images, while the focal length was allowed to vary. Since no zooming was used, any differences in the values of the focal lengths are due either to focusing, which should produce only a marginal effect, or to inaccuracies of the self-calibration algorithm. The focal lengths found for the five cameras, expressed in pixels, were 1372.1, 1354.1, 1390.4, 1361.5, and 1363.8. The standard deviation of these values is 13.8, or about 1% of their mean, showing the stability of the result.

Once the cameras were calibrated, a 3D model was built using the reconstruction package *PhotoBuilder* [129]. The final model is shown in figure 5.14. It is important to notice that, although the relative motion of each pair of cameras is nearly critical, since all the cameras are approximately pointing at the same point in the corner of the building, the parameters were accurately estimated as demonstrated by the visual quality of the reconstruction.

5.7 Conclusions

This chapter presented a novel self-calibration technique based on the Huang and Faugeras constraints on essential matrices. This constraint was fully exploited, providing (i) a linear algorithm for computing focal lengths that generalises for an arbitrary number of cameras the results in [61, 13], (ii) a built-in method for the detection of critical motions for each pair of images in the sequence, and (iii) a nonlinear technique for refining the initial estimate of the focal lengths and computing the principal point of each camera. The algorithm was tested with both synthetic and real data, showing good robustness to noise. A Matlab implementation



Figure 5.13. First, second, fourth and fifth images in the outdoor sequence used for reconstruction.

of the complete self-calibration algorithm is publicly available in <ftp://svr-ftp.eng.cam.ac.uk/pub/>.

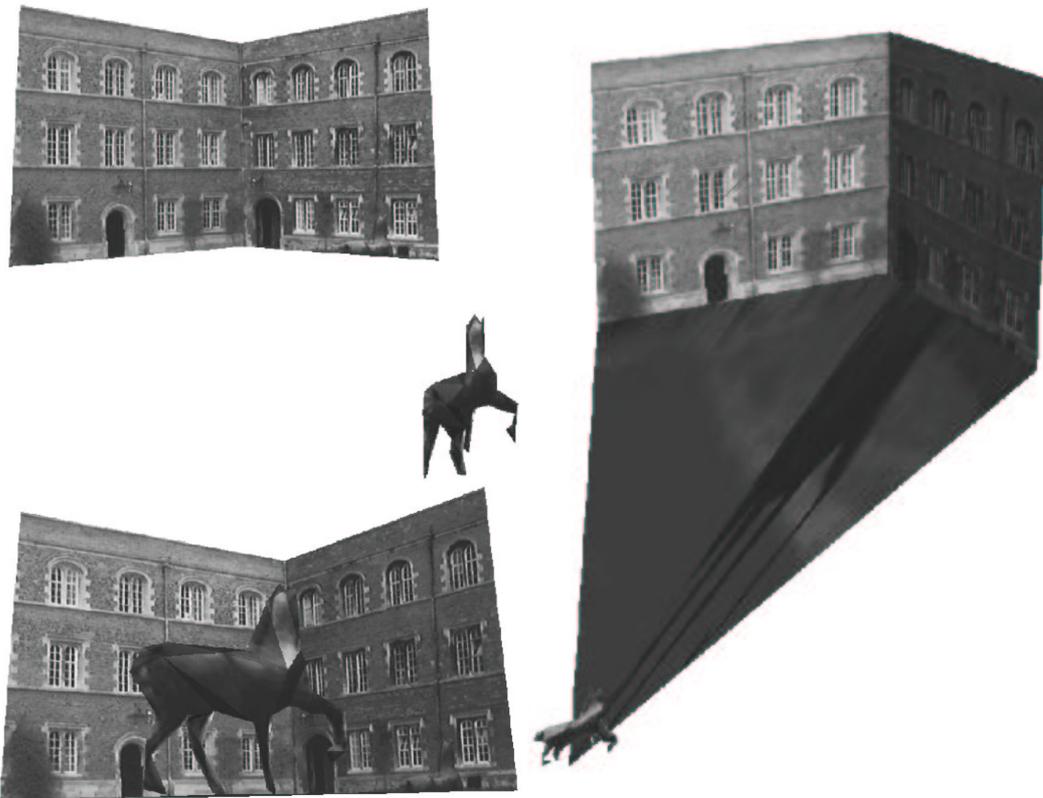


Figure 5.14. 3D model reconstructed from the outdoor sequence shown at different view-points.

Chapter 6

Conclusion

6.1 Summary

This thesis attempted to contribute to the understanding of two important problems in computer vision: estimation of epipolar and camera self-calibration. By considering an important subclass of camera motions (circular motion), it provided the first practical solution to the estimation of epipolar geometry based solely on profiles. A different standpoint was taken for the problem of self-calibration, which was considered in its most general form (varying intrinsic parameters) and was tackled in conjunction with the analysis of critical motions.

Estimation of Epipolar Geometry from Profiles. The central idea that allowed the solution of the problem of estimating the epipolar geometry of a turntable sequence from profiles was the recognition that, as the object placed on the turntable rotates, it sweeps out a surface of revolution. Symmetry properties of the profile of this surface of revolution provided important components of the epipolar geometry of such image sequences, i.e., the image of the axis of rotation, and a special van-

ishing point, the pole, which corresponds to the image of the point at infinity in the direction orthogonal to the plane containing the optical centre of the virtual camera that sees the surface of revolution and the axis of the turntable.

The remaining components of the epipolar geometry are the horizon, which is the image of the plane containing the camera centres, and the position of one epipole along the horizon. These are obtained by exploiting the epipolar constraint for each pair of profiles, which provided an initial estimate for the epipoles of many pairs of images. The horizon was then computed by robustly fitting a line to the cloud of epipoles, and these were then re-estimated from the epipolar constraint, but now with the additional constraint that they should lie on the horizon.

The accuracy of the epipolar geometry so obtained was verified by using it together with the camera intrinsic parameters to estimate the camera motion and check it against available ground truths, producing excellent results. Furthermore, some experiments of reconstruction from profiles were carried out to demonstrate the quality of the algorithms developed here.

Camera Self-Calibration. The rigid body constraint is the principle behind self-calibration algorithms. It has many algebraic interpretations, such as the Kruppa equations, the Trivedi constraints, and the Huang and Faugeras constraints, which were the ones used in this work. The Huang and Faugeras constraints can be expressed in different forms, and one of them can be used to determine the null space of the essential matrix, which can be easily related to null space of the fundamental matrix, i.e., the epipole. This approach provided a linear solution for the estimation of focal lengths obtained by assuming that the skew, aspect ratio and principal point of a set of cameras are known.

The stability of the solution of the linear system (of type $\mathbf{Ax} = \mathbf{b}$) used in the

computation of the focal lengths of a pair of cameras related to a given fundamental matrix was used as an indicator of how close to critical the relative motion between the cameras was. The measure of stability chosen was the condition number of the matrix \mathbf{A} , and the system of equations was then extended by considering all the fundamental matrices that are related to pairs of cameras that include a fixed camera \mathbf{P} , providing two equations per fundamental matrix for the computation of the focal length of \mathbf{P} . Each new pair of equations of the extended system was weighted by the condition number obtained from its corresponding fundamental matrix.

The next step was then a nonlinear refinement of the estimated focal lengths, which also allowed for the computation of other intrinsic parameters, in particular the principal point. Another formulation of the Huang and Faugeras constraints was adopted, and the condition numbers used in the linear algorithm were again employed to weight the contribution of each fundamental matrix in the nonlinear refinement.

Several experiments with both synthetic and real data were performed, demonstrating the accuracy and robustness of the method in different configurations of fixed, fixed but unknown, and finally varying intrinsic parameters.

6.2 Future Work

It is hoped that this work has presented relevant contributions to the subjects it investigated. However, it is clear that there are many important questions it has not touched:

- the overlooking of the information provided by profiles as a clue for camera motion has been severely criticised in this work; however, that does not mean

that corners and edges should be completely ignored, but, instead, used in conjunction with profiles; a framework to efficiently integrate the information offered by these different visual cues is a topic of current research;

- so far, there is no *optimal* solution to the problem of structure and motion from profiles — there is no equivalent, in this context, of a bundle adjustment algorithm; an important step in filling this gap is the development of a model for the effect that image noise has on the extraction of contours; only when such an error model becomes available can a *maximum likelihood estimator* for structure and motion from profiles be developed;
- the use of the condition numbers of the matrices computed in the linear estimation of the focal lengths as presented here, although useful, is only an *ad hoc* measure of how far to critical the camera motion is; any self-calibration procedure would benefit from a more thorough investigation of this problem.

Appendix A

Derivatives of the Geometric Error of the Fundamental Matrix

The minimisation of geometric error (3.24) in the computation of the fundamental matrix is usually carried out through numerical techniques [99, 126]. However, these techniques involve the computation of the gradient and sometimes the Hessian of the cost function, which are normally estimated by using finite differences [28, pp. 154]. The computation of the Hessian of (3.24) serves the additional purpose of providing error bounds to the accuracy of the fundamental matrix being estimated through covariance propagation [35, 164]. However, finite difference methods for computing derivatives can be inaccurate unless high order approximations are made [52], which can be computationally expensive. This appendix presents analytical expressions for both the gradient and the Hessian of (3.24), which can be used to overcome these difficulties.

As in section 3.1.4, we will follow the notation of [104] and define $\text{vec}(\mathbf{X})$ as the vector built from the $m \times n$ matrix \mathbf{X} by stacking its columns in order, from top to bottom, and define \mathbf{C}_{mn} as the commutation matrix (m, n) , i.e., $\text{vec}(\mathbf{X}^T) =$

$\mathbf{C}_{mn} \text{vec}(\mathbf{X})$. The element c_{pq} of \mathbf{C}_{mn} is given by

$$c_{pq} = \begin{cases} 1 & \text{if } q = m((p-1) \bmod m) + \lfloor (p-1)/m \rfloor + 1, \\ 0 & \text{otherwise.} \end{cases} \quad (\text{A.1})$$

The gradient of a vector-valued function $f = f(\mathbf{x})$ is a row vector, represented as $\frac{df}{d\mathbf{x}}$. Finally, let

$$\mathbf{D} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \quad (\text{A.2})$$

A.1 Gradient of the Geometric Error

Let

$$\mathfrak{r}_i = \mathbf{x}_i^T \mathbf{F}^T \mathbf{D} \mathbf{F} \mathbf{x}_i \quad (\text{A.3})$$

$$\mathfrak{r}'_i = \mathbf{x}'_i{}^T \mathbf{F} \mathbf{D} \mathbf{F}^T \mathbf{x}'_i. \quad (\text{A.4})$$

The derivative of the geometric error (3.24) with respect to \mathbf{F} , $d\epsilon_i^2/d\mathbf{F}$, is given by

$$\frac{d\epsilon_i^2}{d\mathbf{F}} = 2\mathbf{x}'_i \mathbf{x}'_i{}^T \mathbf{F} \mathbf{x} \mathbf{x}_i^T \left(\frac{1}{\mathfrak{r}'_i} + \frac{1}{\mathfrak{r}_i} \right) - 2(\mathbf{x}'_i{}^T \mathbf{F} \mathbf{x}_i)^2 \left(\frac{\mathbf{x}'_i \mathbf{x}'_i{}^T \mathbf{F} \mathbf{D}}{\mathfrak{r}'_i{}^2} + \frac{\mathbf{D} \mathbf{F} \mathbf{x}_i \mathbf{x}_i^T}{\mathfrak{r}_i^2} \right). \quad (\text{A.5})$$

Consider a fundamental matrix \mathbf{F} parameterised as

$$\mathbf{F} = \begin{bmatrix} x & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & 1 \end{bmatrix}, \quad (\text{A.6})$$

where x is computed from f_{ij} so that $\det(\mathbf{F}) = 0$, i.e., $x = n/d$, where

$$n = f_{31}f_{13}f_{22} + f_{21}f_{12} - f_{21}f_{13}f_{32} - f_{31}f_{12}f_{23} \quad (\text{A.7})$$

$$d = f_{22} - f_{23}f_{32} \quad (\text{A.8})$$

It can be argued that this parameterisation is not general, since it precludes the point $\mathbf{x} = [0 \ 0 \ 1]^T$ and $\mathbf{x}' = \mathbf{x}$ from being related by the epipolar constraint, since $\mathbf{x}'^T \mathbf{F} \mathbf{x} = 1 \neq 0$. In practice, however, this situation is unlikely to occur, and even if it does a simple translation in the coordinate system of the images will suffice to correct the situation. Moreover, the parameterisation is minimal, since it involves only seven parameters. Using this parameterisation and adopting the notation $\mathbf{f} = [f_{12} \ f_{13} \ f_{21} \ f_{22} \ f_{23} \ f_{31} \ f_{32}]^T$, the tensor $\frac{d\mathbf{F}}{d\mathbf{f}}$ can be easily computed:

$$\frac{d\mathbf{F}}{df_{12}} = \begin{bmatrix} \frac{f_{21} - f_{23}f_{31}}{d} & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad (\text{A.9})$$

$$\frac{d\mathbf{F}}{df_{13}} = \begin{bmatrix} \frac{f_{31}f_{22} - f_{21}f_{32}}{d} & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad (\text{A.10})$$

$$\frac{d\mathbf{F}}{df_{21}} = \begin{bmatrix} \frac{f_{12} - f_{13}f_{32}}{d} & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad (\text{A.11})$$

$$\frac{d\mathbf{F}}{df_{22}} = \begin{bmatrix} \frac{f_{13}f_{31}d - n}{d^2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad (\text{A.12})$$

$$\frac{d\mathbf{F}}{df_{23}} = \begin{bmatrix} \frac{f_{32}n - f_{12}f_{31}\vartheta}{\vartheta^2} & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad (\text{A.13})$$

$$\frac{d\mathbf{F}}{df_{31}} = \begin{bmatrix} \frac{f_{13}f_{22} - f_{12}f_{23}}{\vartheta} & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad (\text{A.14})$$

$$\frac{d\mathbf{F}}{df_{32}} = \begin{bmatrix} \frac{f_{23}n - f_{13}f_{21}\vartheta}{\vartheta^2} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}. \quad (\text{A.15})$$

and, at last,

$$\frac{d\epsilon_i^2}{df_{ij}} = \left(\text{vec} \left(\frac{d\epsilon_i^2}{d\mathbf{F}} \right) \right)^T \text{vec} \left(\frac{d\mathbf{F}}{df_{ij}} \right) \quad (\text{A.16})$$

A.2 Hessian of the Geometric Error

To compute the Hessian $\frac{d^2\epsilon_i}{d \text{vec } \mathbf{F} d \text{vec }^T \mathbf{F}}$, observe that

$$\frac{d\mathbf{x}_i}{d \text{vec } \mathbf{F}} = 2\mathbf{x}_i^T \otimes (\mathbf{D}\mathbf{F}\mathbf{x}_i)^T \quad (\text{A.17})$$

and, therefore,

$$\frac{d^2\mathbf{x}_i}{d \text{vec } \mathbf{F} d \text{vec }^T \mathbf{F}} = 2(\mathbf{x}_i \otimes \mathbf{x}_i)^T \otimes \text{vec } \mathbf{D}. \quad (\text{A.18})$$

Analogous expressions can be found for $\frac{dx'_i}{d \text{vec } \mathbf{F}}$ and $\frac{d^2 x'_i}{d \text{vec } \mathbf{F} d \text{vec}^T \mathbf{F}}$. Moreover,

$$\frac{d(\mathbf{x}'_i{}^T \mathbf{F} \mathbf{x}_i)^2}{d \text{vec } \mathbf{F}} = 2\mathbf{x}'_i{}^T \mathbf{F} \mathbf{x}_i \mathbf{x}_i{}^T \otimes \mathbf{x}'_i{}^T \quad (\text{A.19})$$

$$\frac{d^2(\mathbf{x}'_i{}^T \mathbf{F} \mathbf{x}_i)^2}{d \text{vec } \mathbf{F} d \text{vec}^T \mathbf{F}} = 2\mathbf{x}_i \otimes \mathbf{x}'_i \otimes \mathbf{x}_i{}^T \otimes \mathbf{x}'_i{}^T. \quad (\text{A.20})$$

The Hessian of the geometric error is then found by substituting (A.17), (A.18) and the corresponding expressions for $\frac{dx'_i}{d \text{vec } \mathbf{F}}$ and $\frac{d^2 x'_i}{d \text{vec } \mathbf{F} d \text{vec}^T \mathbf{F}}$, together with (A.19) and (A.20), into

$$\begin{aligned} \frac{d^2 \epsilon_i}{d \text{vec } \mathbf{F} d \text{vec}^T \mathbf{F}} &= \left(\frac{1}{x_i} + \frac{1}{x'_i} \right) \frac{d^2(\mathbf{x}'_i{}^T \mathbf{F} \mathbf{x}_i)^2}{d \text{vec } \mathbf{F} d \text{vec}^T \mathbf{F}} + (\mathbf{x}'_i{}^T \mathbf{F} \mathbf{x}_i)^2 \times \\ &\quad \left(2 \frac{\left(\frac{dx_i}{d \text{vec } \mathbf{F}} \right)^T \frac{dx_i}{d \text{vec } \mathbf{F}}}{x_i^3} - \frac{d^2 x_i}{d \text{vec } \mathbf{F} d \text{vec}^T \mathbf{F}} \frac{1}{x_i^2} + 2 \frac{\left(\frac{dx'_i}{d \text{vec } \mathbf{F}} \right)^T \frac{dx'_i}{d \text{vec } \mathbf{F}}}{x_i'^3} - \frac{d^2 x'_i}{d \text{vec } \mathbf{F} d \text{vec}^T \mathbf{F}} \frac{1}{x_i'^2} \right) \\ &- \left(\frac{\left(\frac{dx_i}{d \text{vec } \mathbf{F}} \right)^T}{x_i^2} + \frac{\left(\frac{dx'_i}{d \text{vec } \mathbf{F}} \right)^T}{x_i'^2} \right) \frac{d(\mathbf{x}'_i{}^T \mathbf{F} \mathbf{x}_i)^2}{d \text{vec } \mathbf{F}} \\ &- \left(\frac{d(\mathbf{x}'_i{}^T \mathbf{F} \mathbf{x}_i)^2}{d \text{vec } \mathbf{F}} \right)^T \left(\frac{dx_i}{d \text{vec } \mathbf{F}} \frac{1}{x_i^2} + \frac{dx'_i}{d \text{vec } \mathbf{F}} \frac{1}{x_i'^2} \right). \end{aligned} \quad (\text{A.21})$$

Appendix B

Derivative of the Huang and Faugeras Constraints

It can be shown that $\text{vec}(\mathbf{XYZ}) = (\mathbf{Z}^T \otimes \mathbf{X}) \text{vec } \mathbf{Y}$ and, if \mathbf{Y} is an $r \times s$ matrix, $\mathbf{C}_{rm}(\mathbf{X} \otimes \mathbf{Y}) = (\mathbf{Y} \otimes \mathbf{X})\mathbf{C}_{sn}$ [104]. Let $f(\sigma_1^{ij}, \sigma_2^{ij}) = 1 - \sigma_2^{ij}/\sigma_1^{ij}$. Then

$$\frac{df}{d \text{vec } \mathbf{K}_i} = \frac{\frac{d\sigma_1^{ij}}{d \text{vec } \mathbf{K}_i} \sigma_2^{ij} - \frac{d\sigma_2^{ij}}{d \text{vec } \mathbf{K}_i} \sigma_1^{ij}}{(\sigma_1^{ij})^2}. \quad (\text{B.1})$$

But

$$\frac{d\sigma_k^{ij}}{d \text{vec } \mathbf{K}_i} = \frac{d\sigma_k^{ij}}{d \text{vec } \mathbf{E}_{ij}} \frac{d \text{vec } \mathbf{E}_{ij}}{d \text{vec } \mathbf{K}_i}, \quad (\text{B.2})$$

$$\frac{d \text{vec } \mathbf{E}_{ij}}{d \text{vec } \mathbf{K}_i} = (\mathbb{I} \otimes (\mathbf{K}_j^T \mathbf{F}_{ij})), \text{ and} \quad (\text{B.3})$$

$$\frac{d\sigma_k^{ij}}{d \text{vec } \mathbf{E}_{ij}} = (\mathbf{v}_k^{ij})^T \otimes (\mathbf{u}_k^{ij})^T \text{ (see[104])}. \quad (\text{B.4})$$

where \mathbf{u}_k^{ij} and \mathbf{v}_k^{ij} are the k -th columns of the matrices \mathbf{U}^{ij} and \mathbf{V}^{ij} in (5.28). Substituting (B.4) and (B.3) in (B.2), and the result in (B.1), one obtains the desired

derivative:

$$\frac{df}{d \text{vec } \mathbf{K}_i} = \frac{\sigma_2^{ij} (\mathbf{v}_1^{ij})^T \otimes ((\mathbf{u}_1^{ij})^T \mathbf{K}_j^T \mathbf{F}_{ij}) - \sigma_1^{ij} (\mathbf{v}_2^{ij})^T \otimes ((\mathbf{u}_2^{ij})^T \mathbf{K}_j^T \mathbf{F}_{ij})}{(\sigma_1^{ij})^2} \quad (\text{B.5})$$

A similar manipulation can be used to derive, with the aid of some properties of the Kronecker product and the commutation matrix, the Jacobian of $f(\sigma_1^{ij}, \sigma_2^{ij})$ with respect to \mathbf{K}_j :

$$\frac{df}{d \text{vec } \mathbf{K}_i} = \frac{\sigma_2^{ij} ((\mathbf{v}_1^{ij})^T \mathbf{K}_i^T \mathbf{F}_{ij}^T) \otimes (\mathbf{u}_1^{ij})^T - \sigma_1^{ij} ((\mathbf{v}_2^{ij})^T \mathbf{K}_i^T \mathbf{F}_{ij}^T) \otimes (\mathbf{u}_2^{ij})^T}{(\sigma_1^{ij})^2} \mathbf{C}_{33}, \quad (\text{B.6})$$

where \mathbf{C}_{33} is the commutation matrix (3, 3) (see (A.1)).

Bibliography

- [1] J. Y. Aloimonos. Perspective approximations. *Image and Vision Computing*, 8(3):179–192, August 1990.
- [2] P. Anandan, K. Hanna, and R. Kumar. Shape recovery from multiple views: A parallax based approach. In *Proc. of 12th Int. Conf. on Pattern Recognition*, pages A:685–688, Seattle, USA, June 1994.
- [3] P. Anandan, M. Irani, P. H. S. Torr, and A. Zisserman. Panel session for direct versus features methods. In B. Triggs, A. Zisserman, and R. Szeliski, editors, *Vision Algorithms: Theory and Practice*, number 1883 in Lecture Notes in Computer Science, pages 295–297, Corfu, Greece, September 1999. Springer-Verlag.
- [4] M. Armstrong, A. Zisserman, and R. Hartley. Self-calibration from image triplets. In B. Buxton and R. Cipolla, editors, *Proc. 4th European Conf. on Computer Vision*, volume I of *Lecture Notes in Computer Science 1064*, pages 3–16, Cambridge, UK, April 1996. Springer-Verlag.
- [5] K. Åström, R. Cipolla, and P. Giblin. Generalised epipolar constraints. *Int. Journal of Computer Vision*, 33(1):51–72, September 1999.

-
- [6] K. Åström, R. Cipolla, and P. J. Giblin. Generalised epipolar constraints. In B. F. Buxton and R. Cipolla, editors, *Proc. 4th European Conf. on Computer Vision*, volume II of *Lecture Notes in Computer Science 1065*, pages 97–108, Cambridge, UK, April 1996. Springer-Verlag.
- [7] H. G. Barrow and J. M. Tenenbaum. Interpreting line drawings as three-dimensional surfaces. *Artificial Intelligence*, 17:75–116, August 1981.
- [8] P. A. Beardsley. *Applications of Projective Geometry to Robot Vision*. PhD thesis, Department of Engineering Science, University of Oxford, Oxford, 1992.
- [9] P. A. Beardsley, P. H. S. Torr, and A. Zisserman. 3D model acquisition from extended image sequences. In B. Buxton and R. Cipolla, editors, *Proc. 4th European Conf. on Computer Vision*, volume II of *Lecture Notes in Computer Science 1065*, pages 683–695, Cambridge, UK, April 1996. Springer-Verlag.
- [10] P. A. Beardsley, A. Zisserman, and D. W. Murray. Sequential updating of projective and affine structure from motion. *Int. Journal of Computer Vision*, 23(3):235–259, 1997.
- [11] A. Blake. Specular stereo. In *Proc. 9th Int. Joint Conference on Artificial Intell.*, pages 973–976, 1985.
- [12] A. Blake and M. Isard. *Active Contours : The Application of Techniques from Graphics, Vision, Control Theory and Statistics to Visual Tracking of Shapes in Motion*. Springer-Verlag, London, 1998.

-
- [13] S. Bougnoux. From projective to Euclidean space under any practical situation, a criticism of self-calibration. In *Proc. 6th Int. Conf. on Computer Vision*, pages 790–796, Bombay, India, January 1998.
- [14] E. Boyer and M. O. Berger. 3D surface reconstruction using occluding contours. *Int. Journal of Computer Vision*, 22(3):219–233, March/April 1997.
- [15] C. G. Broyden. The convergence of a class of double rank minimization algorithms: Part I. *J. Inst. Maths. Applns.*, 6:76–90, 1970.
- [16] J. F. Canny. Finding edges and lines in images. Master’s thesis, MIT, Cambridge, USA, 1983.
- [17] J. F. Canny. A computational approach to edge detection. *IEEE Trans. Pattern Analysis and Machine Intell.*, 8(6):679–698, November 1986.
- [18] N. Canterakis. A minimal set of constraints for the trifocal tensor. In D. Vernon, editor, *Proc. 6th European Conf. on Computer Vision*, volume I of *Lecture Notes in Computer Science 1842*, pages 84–99, Dublin, Ireland, June/July 2000. Springer-Verlag.
- [19] S. Carlsson and D. Weinshall. Dual computation of projective shape and camera positions from multiple images. *Int. Journal of Computer Vision*, 27(3):227–241, May 1998.
- [20] T. J. Cham and R. Cipolla. Geometric saliency of curve correspondences and grouping of symmetric contours. In B. F. Buxton and R. Cipolla, editors, *Proc. 4th European Conf. on Computer Vision*, volume I of *Lecture Notes in Computer Science 1064*, pages 385–398, Cambridge, UK, April 1996. Springer-Verlag.

-
- [21] T. J. Cham and R. Cipolla. MDL-based curve representation using b-spline active contours. In R. B. Fisher, editor, *Proc. British Machine Vision Conference*, pages 363–372, Edinburgh, UK, September 1996.
- [22] S. Chaudhuri and A. N. Rajagopalan. *Depth from Defocus: A Real Aperture Imaging Approach*. Springer-Verlag, January 1999.
- [23] C.-T. Chen. *Linear Systems Theory and Design*. Holt, Rinehart and Winston, New York, 1984.
- [24] R. Cipolla. The visual motion of curves and surfaces. *Phil. Trans. Royal Soc. London A*, 356(1740):1103–1121, May 1998.
- [25] R. Cipolla, K. Åström, and P. J. Giblin. Motion from the frontier of curved surfaces. In *Proc. 5th Int. Conf. on Computer Vision*, pages 269–275, Cambridge, USA, June 1995.
- [26] R. Cipolla and A. Blake. The dynamic analysis of apparent contours. In *Proc. 3rd Int. Conf. on Computer Vision*, pages 616–623, Osaka, Japan, December 1990.
- [27] R. Cipolla and A. Blake. Surface shape from the deformation of apparent contours. *Int. Journal of Computer Vision*, 9(2):83–112, November 1992.
- [28] R. Cipolla and P. J. Giblin. *Visual Motion of Curves and Surfaces*. Cambridge University Press, Cambridge, UK, 1999.
- [29] R. Cipolla, Y. Okamoto, and Y. Kuno. Robust structure from motion using motion parallax. In *Proc. 4th Int. Conf. on Computer Vision*, pages 374–382, Berlin, Germany, May 1993.

- [30] H. S. M. Coxeter. *Introduction to Geometry*. John Wiley and Sons, New York, 2nd edition, 1969.
- [31] H. S. M. Coxeter. *Projective Geometry*. Springer-Verlag, New York, 1994 reprint of 2nd edition, 1974.
- [32] A. Criminisi, I. Reid, and A. Zisserman. Single view metrology. In *Proc. 7th Int. Conf. on Computer Vision*, volume I, pages 434–441, Corfu, Greece, September 1999.
- [33] G. Cross, A. Fitzgibbon, and A. Zisserman. Parallax geometry of smooth surfaces in multiple views. In *Proc. 7th Int. Conf. on Computer Vision*, volume I, pages 323–329, Corfu, Greece, September 1999.
- [34] G. Cross and A. Zisserman. Quadric reconstruction from dual-space geometry. In *Proc. 6th Int. Conf. on Computer Vision*, pages 25–31, Bombay, India, January 1998.
- [35] G. Csurka, Zeller C., and Zhang Z. Faugeras O. D. Characterizing the uncertainty of the fundamental matrix. *Computer Vision and Image Understanding*, 68(1):18–36, October 1997.
- [36] R. Curwen and A. Blake. Dynamic contours: Real-time active splines. In A. Blake and A. Yuille, editors, *Active Vision*, Artificial Intelligence Series, chapter 3, pages 39–57. MIT Press, Cambridge, USA, December 1992.
- [37] R. W. Curwen, C. V. Stewart, and J. L. Mundy. Recognition of plane projective symmetry. In *Proc. 6th Int. Conf. on Computer Vision*, pages 1115–1122, Bombay, India, January 1998.

-
- [38] R. Deriche. Using Canny's criteria to derive a recursively implemented optimal edge detector. *Int. Journal of Computer Vision*, 1(2):167–187, May 1987.
- [39] J. A. Eisele and R. M. Mason. *Applied Tensor and Matrix Analysis*. John Wiley & Sons, New York, 1970.
- [40] O. D. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig. In G. Sandini, editor, *Proc. 2nd European Conf. on Computer Vision*, Lecture Notes in Computer Science 588, pages 563–578. Springer-Verlag, May 1992.
- [41] O. D. Faugeras. *Three-Dimensional Computer Vision: a Geometric Viewpoint*. Artificial Intelligence Series. MIT Press, Cambridge, USA, 1993.
- [42] O. D. Faugeras. Stratification of three-dimensional vision: Projective, affine and metric representations. *J. Opt. Soc. America A*, 12(3):465–484, March 1995.
- [43] O. D. Faugeras and R. Keriven. Complete dense stereovision using level set methods. In H. Burkhardt and B. Neumann, editors, *Proc. 5th European Conf. on Computer Vision*, volume I of *Lecture Notes in Computer Science 1406*, pages 379–393, Freiburg, Germany, June 1998. Springer-Verlag.
- [44] O. D. Faugeras, Q. T. Luong, and S. J. Maybank. Camera self-calibration: Theory and experiments. In G. Sandini, editor, *Proc. 2nd European Conf. on Computer Vision*, Lecture Notes in Computer Science 588, pages 321–334. Springer-Verlag, May 1992.

- [45] O. D. Faugeras and B. Mourrain. On the geometry and algebra of the point and line correspondences between N images. In *Proc. 5th Int. Conf. on Computer Vision*, pages 951–962, Cambridge, USA, June 1995.
- [46] O. D. Faugeras and T. Papadopoulo. Grassman-Cayley algebra for modelling systems of cameras and the algebraic equations of the manifold of trifocal tensors. *Phil. Trans. Royal Soc. London A*, 356(1740):1123–1152, May 1998.
- [47] O. D. Faugeras and T. Papadopoulo. A nonlinear method for estimating the projective geometry of 3 views. In *Proc. 6th Int. Conf. on Computer Vision*, pages 477–484, Bombay, India, January 1998.
- [48] O. D. Faugeras and L. Robert. What can two images tell us about a third one? *Int. Journal of Computer Vision*, 18(1):5–19, April 1996.
- [49] O. D. Faugeras and G. Toscani. The calibration problem for stereo. In *Proc. Conf. Computer Vision and Pattern Recognition*, pages 15–20, Miami, USA, June 1986.
- [50] A. W. Fitzgibbon, G. Cross, and A. Zisserman. Automatic 3D model construction for turn-table sequences. In R. Koch and L. Van Gool, editors, *3D Structure from Multiple Images of Large-Scale Environments, European Workshop SMILE'98*, Lecture Notes in Computer Science 1506, pages 155–170, Freiburg, Germany, June 1998. Springer-Verlag.
- [51] R. Fletcher. A new approach to variable metric algorithms. *Computer J.*, 13(13):317–322, 1970.
- [52] B. H. Flowers. *An Introduction to Numerical Methods in C++*. Clarendon Press, Oxford, UK, 1996.

- [53] D. Forsyth. The outline of an algebraic surface yields the surface. In R. B. Fisher, editor, *Design and Application of Curves and Surfaces — Mathematics of Surfaces V*, number 50 in The Institute of Mathematics and its Applications Series, pages 413–423, Edinburgh, UK, 1994. Clarendon Press.
- [54] P. J. Giblin, F. E. Pollick, and J. E. Rycroft. Recovery of an unknown axis of rotation from the profiles of a rotating surface. *J. Opt. Soc. America A*, 11(7):1976–1984, July 1994.
- [55] P. J. Giblin and R. S. Weiss. Reconstruction of surfaces from profiles. In *Proc. 1st Int. Conf. on Computer Vision*, pages 136–144, London, June 1987.
- [56] P. J. Giblin and R. S. Weiss. Epipolar fields on surfaces. In J.-O. Eklundh, editor, *Proc. 3rd European Conf. on Computer Vision*, volume I of *Lecture Notes in Computer Science 800*, pages 14–23, Stockholm, Sweden, May 1994. Springer-Verlag.
- [57] D. Goldfarb. A family of variable metric methods derived by variational means. *Maths. Comput.*, 24:23–26, 1970.
- [58] G. H. Golub and C. van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, USA, 3rd edition, 1996.
- [59] C. G. Harris. Geometry from visual motion. In A. Blake and A. Yuille, editors, *Active Vision*, pages 263–284. MIT Press, Cambridge, USA, 1992.
- [60] C. G. Harris and M. Stephens. A combined corner and edge detector. In *4th Alvey Vision Conference*, pages 147–152, Manchester, UK, August/September 1988.

- [61] R. Hartley. Estimation of relative camera positions for uncalibrated cameras. In G. Sandini, editor, *Proc. 2nd European Conf. on Computer Vision*, Lecture Notes in Computer Science 588, pages 579–587, Santa Margherita Ligure, Italy, May 1992. Springer–Verlag.
- [62] R. Hartley. Euclidean reconstruction from uncalibrated views. In J. L. Mundy, A. Zisserman, and D. Forsyth, editors, *Applications of Invariance in Computer Vision*, Lecture Notes in Computer Science 825, pages 237–256. Springer-Verlag, 1994.
- [63] R. Hartley. Projective reconstruction and invariants from multiple images. *IEEE Trans. Pattern Analysis and Machine Intell.*, 16(10):1036–1041, October 1994.
- [64] R. I. Hartley. A linear method for the reconstruction of lines and points. In *Proc. 5th Int. Conf. on Computer Vision*, pages 882–887, Cambridge, USA, June 1995.
- [65] R. I. Hartley. In defense of the eight-point algorithm. *IEEE Trans. Pattern Analysis and Machine Intell.*, 19(6):580–593, June 1997.
- [66] R. I. Hartley. Lines and points in three views and the trifocal tensor. *Int. Journal of Computer Vision*, 22(2):125–140, March 1997.
- [67] R. I. Hartley. Chirality. *Int. Journal of Computer Vision*, 26(1):41–61, January 1998.
- [68] R. I. Hartley. Minimising algebraic error in geometric estimation problems. In *Proc. 6th Int. Conf. on Computer Vision*, pages 469–476, Bombay, India, January 1998.

-
- [69] R. I. Hartley and P. Sturm. Triangulation. *Computer Vision and Image Understanding*, 68(2):146–157, November 1997.
- [70] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge, UK, 2000.
- [71] A. Heyden. A common framework for multiple view tensors. In H. Burkhardt and B. Neumann, editors, *Proc. 5th European Conf. on Computer Vision*, volume I of *Lecture Notes in Computer Science 1406*, pages 3–19, Freiburg, Germany, June 1998. Springer-Verlag.
- [72] A. Heyden, R. Berthilsson, and G. Sparr. An iterative factorization method for projective structure and motion from image sequences. *Image and Vision Computing*, 17(13):981–991, November 1999.
- [73] R. Horaud, F. Dornaika, B. Lamiroy, and S. Christy. Object pose: The link between weak perspective, paraperspective, and full perspective. *Int. Journal of Computer Vision*, 22(2):173–189, March 1997.
- [74] B. K. P. Horn. *Robot Vision*. McGraw-Hill, New York, 1986.
- [75] B. K. P. Horn and M. J. Brooks, editors. *Shape from Shading*. Artificial Intelligence Series. MIT Press, Massachusetts, USA, 1989.
- [76] T. S. Huang and O. Faugeras. Some properties of the e matrix in two-view motion estimation. *IEEE Trans. Pattern Analysis and Machine Intell.*, 11(12):1310–1312, December 1989.
- [77] M. Irani and P. Anandan. About direct methods. In B. Triggs, A. Zisserman, and R. Szeliski, editors, *Vision Algorithms: Theory and Practice*, number

- 1883 in *Lecture Notes in Computer Science*, pages 267–277, Corfu, Greece, September 1999. Springer-Verlag.
- [78] M. Irani, P. Anandan, and D. Weinshall. From reference frames to reference planes: Multi-view parallax geometry and applications. In H. Burkhardt and B. Neumann, editors, *Proc. 5th European Conf. on Computer Vision*, volume II of *Lecture Notes in Computer Science 1407*, pages 829–845, Freiburg, Germany, June 1998. Springer-Verlag.
- [79] T. Joshi, N. Ahuja, and J. Ponce. Structure and motion estimation from dynamic silhouettes under perspective projection. *Int. Journal of Computer Vision*, 31(1):31–50, February 1999.
- [80] F. Kahl, B. Triggs, and K. Åström. Critical motions for auto-calibration when some intrinsic parameters can vary. *Journal of Mathematical Image and Vision*, 13(2):131–146, October 2000.
- [81] T. Kanade and J. R. Kender. Mapping image properties into shape constraints: Skewed symmetry, affine-transformable patterns, and the shape-from-texture paradigm. In J. Beck, B. Hope, and A. Rosenfeld, editors, *Human and Machine Vision*, pages 237–257. Academic Press, New York, 1983.
- [82] T. Kanade and M. Okutomi. A stereo matching algorithm with an adaptive window: Theory and experiment. In *Proc. of the IEEE Intl. Conf. on Robotics and Automation*, pages 1088–1095, Sacramento, USA, April 1991.
- [83] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. In *Proc. 1st Int. Conf. on Computer Vision*, pages 259–268, London, June 1987.

-
- [84] R. Koch, M. Pollefeys, and L. Van Gool. Multi viewpoint stereo from uncalibrated video sequences. In H. Burkhardt and B. Neumann, editors, *Proc. 5th European Conf. on Computer Vision*, volume I of *Lecture Notes in Computer Science 1406*, pages 55–71, Freiburg, Germany, June 1998. Springer-Verlag.
- [85] J. J. Koenderink. What does the occluding contour tell us about solid shape? *Perception*, 13:321–330, 1984.
- [86] J. J. Koenderink. Optic flow. *Vision Research*, 26(1):161–179, 1986.
- [87] J. J. Koenderink. *Solid Shape*. MIT Press, Cambridge, USA, 1990.
- [88] J. J. Koenderink and A. J. Van Doorn. Photometric invariants related to solid shape. *Optica Acta*, 27(7):981–996, 1980.
- [89] K. N. Kutulakos and S. M. Seitz. A theory of shape by space carving. In *Proc. 7th Int. Conf. on Computer Vision*, volume I, pages 307–314, Corfu, Greece, September 1999.
- [90] S. Lang. *Algebra*. Addison-Wesley, Menlo Park, California, 2nd edition, 1984.
- [91] J. Lasenby and A. N. Lasenby. Estimating tensors for matching over multiple views. *Phil. Trans. Royal Soc. London A*, 356(1740):1267–1282, May 1998.
- [92] J. M. Lawn and R. Cipolla. Robust egomotion estimation from affine motion-parallax. In J.-O. Eklundh, editor, *Proc. 3rd European Conf. on Computer Vision*, volume I of *Lecture Notes in Computer Science 800*, pages 205–210, Stockholm, Sweden, May 1994. Springer-Verlag.

-
- [93] K. Levenberg. A method for the solution of certain non-linear problems in least squares. *Quart. J. Appl. Maths.*, II(2):164–168, 1944.
- [94] M. D. Levine, D. A. O’Handley, and G. M. Yagi. Computer determination of depth maps. *Computer Graphics and Image Processing*, 2:131–150, 1973.
- [95] J. Liu, J. Mundy, D. Forsyth, A. Zisserman, and C. Rothwell. Efficient recognition of rotationally symmetric surfaces and straight homogeneous generalized cylinders. In *Proc. Conf. Computer Vision and Pattern Recognition*, pages 123–129, New York, June 1993.
- [96] H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, September 1981.
- [97] M. A. Lourakis and R. Deriche. Camera self-calibration using the singular value decomposition of the fundamental matrix: From point correspondences to 3D measurements. Technical Report 3748, INRIA, Sophia Antipolis, France, August 1999.
- [98] M. A. Lourakis and R. Deriche. Camera self-calibration using the singular value decomposition of the fundamental matrix. In *4th Asian Conference on Computer Vision*, volume I, pages 403–408, Taipei, Taiwan, January 2000.
- [99] D. G. Luenberger. *Linear and Nonlinear Programming*. Addison-Wesley, USA, 2nd edition, 1984.
- [100] Q.-T. Luong. *Matrice Fundamentale et Auto-Calibration en Vision Par Ordinateur*. PhD thesis, Universite de Paris-Sud, Orsay, France, 1992.

-
- [101] Q.-T. Luong and O. Faugeras. Self-calibration of a moving camera from point correspondences and fundamental matrices. *Int. Journal of Computer Vision*, 22(3):261–289, 1997.
- [102] Q.-T. Luong and O. D. Faugeras. The fundamental matrix: Theory, algorithms, and stability analysis. *Int. Journal of Computer Vision*, 17(1):43–75, January 1996.
- [103] F. S. Macaulay. *The Algebraic Theory of Modular Systems*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, UK, 1994. Reissue of the 1916 edition.
- [104] J. R. Magnus and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York, 1995.
- [105] J. Malik and R. Rosenholtz. Computing local surface orientation and shape from texture for curved surfaces. *Int. Journal of Computer Vision*, 23(2):149–168, June 1997.
- [106] D. W. Marquardt. An algorithm for least-squares estimation of non-linear parameters. *J. Soc. Ind. Appl. Maths.*, 11(2):431–441, 1963.
- [107] D. Marr and E. Hildreth. Theory of edge detection. *Proc. Roy. Soc. London. B.*, 207:187–217, 1980.
- [108] E. A. Maxwell. *The Methods of Plane Projective Geometry Based on the Use of General Homogeneous Coordinates*. Cambridge University Press, Cambridge, 1964 reprint of 2nd edition, 1946.

- [109] S. Maybank and O. D. Faugeras. A theory of self-calibration of a moving camera. *Int. Journal of Computer Vision*, 8(2):123–151, August 1992.
- [110] P. R. S. Mendonça and R. Cipolla. Analysis and computation of an affine trifocal tensor. In H. P. Lewis and S. M. Nixon, editors, *Proc. British Machine Vision Conference*, volume 1, pages 125–133, Southampton, UK, 1998.
- [111] P. R. S. Mendonça and R. Cipolla. Estimation of epipolar geometry from apparent contours: Affine and circular motion cases. In *Proc. Conf. Computer Vision and Pattern Recognition*, volume I, pages 9–14, Fort Collins, Colorado, June 1999.
- [112] P. R. S. Mendonça and R. Cipolla. A simple technique for self-calibration. In *Proc. Conf. Computer Vision and Pattern Recognition*, volume I, pages 500–505, Fort Collins, Colorado, June 1999.
- [113] P. R. S. Mendonça, K.-Y. K. Wong, and R. Cipolla. Camera pose estimation and reconstruction from image profiles under circular motion. In David Vernon, editor, *Proc. 6th European Conf. on Computer Vision*, volume II of *Lecture Notes in Computer Science 1842*, pages 864–877, Dublin, Ireland, June/July 2000. Springer–Verlag.
- [114] D. P. Mukherjee, A. Zisserman, and J. M. Brady. Shape from symmetry—detecting and exploiting symmetry in affine images. In *Phil. Trans. Royal Soc. London A*, volume 351, pages 77–106, 1995.
- [115] J. L. Mundy and A. Zissermann. *Geometric Invariance in Computer Vision*. Artificial Intelligence Series. MIT Press, Cambridge, USA, 1992.

-
- [116] V. S. Nalwa. Line-drawing interpretation: Bilateral symmetry. *IEEE Trans. Pattern Analysis and Machine Intell.*, 11(10):1117–1120, October 1989.
- [117] S. K. Nayar, W. Watanabe, and M. Noguchi. Real-time focus range sensor. In *Proc. 5th Int. Conf. on Computer Vision*, pages 995–1001, Cambridge, USA, June 1995.
- [118] B. O’Neill. *Elementary Differential Geometry*. Academic Press, New York, 1966.
- [119] M. Oren and S. K. Nayar. A theory of specular surface geometry. *Int. Journal of Computer Vision*, 24(2):105–124, September 1996.
- [120] A. P. Pentland. A new sense for depth of field. *IEEE Trans. Pattern Analysis and Machine Intell.*, 9(4):523–531, July 1987.
- [121] M. Pollefeys, R. Khoch, and L. Van Gool. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. In *Proc. 6th Int. Conf. on Computer Vision*, pages 90–95, Bombay, India, January 1998.
- [122] M. Pollefeys, R. Koch, and L. Van Gool. Self-calibration and metric reconstruction in spite of varying and unknown intrinsic camera parameters. *Int. Journal of Computer Vision*, 32(1):7–25, August 1999.
- [123] J. Ponce, A. Hoogs, and D. J. Kriegman. On using CAD models to compute the pose of curved 3D objects. *Computer Vision, Graphics and Image Processing*, 55(2):184–197, March 1992.
- [124] J. Ponce and D. J. Kriegman. Toward 3D curved object recognition from image contours. In J. L. Mundy and A. Zisserman, editors, *Geometric In-*

- variance in Computer Vision*, Artificial Intelligence Series, chapter 21, pages 408–439. MIT Press, Cambridge, USA, 1992.
- [125] J. Porrill and S. B. Pollard. Curve matching and stereo calibration. *Image and Vision Computing*, 9(1):45–50, February 1991.
- [126] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, Cambridge, UK, 2nd edition, 1992.
- [127] L. Quan. Invariants of six points and projective reconstruction from three uncalibrated images. *IEEE Trans. Pattern Analysis and Machine Intell.*, 17(1):34–46, January 1995.
- [128] J. H. Rieger. Three dimensional motion from fixed points of a deforming profile curve. *Optics Letters*, 11(3):123–125, March 1986.
- [129] D. Robertson and R. Cipolla. An interactive system for constraint-based modelling. In M. Mirmehdi and B. Thomas, editors, *Proc. British Machine Vision Conference*, volume II, pages 536–545, Bristol, UK, 2000.
- [130] J. Sato and R. Cipolla. Affine reconstruction of curved surfaces from uncalibrated views of apparent contours. *IEEE Trans. Pattern Analysis and Machine Intell.*, 21(11):1188–1198, November 1999.
- [131] T. W. Sederberg. Piecewise algebraic surface patches. *Computer Aided Geometric Design*, 2:53–59, 1985.
- [132] T. W. Sederberg and J. P. Snively. Parametrization of cubic algebraic surfaces. In R. R. Martin, editor, *Design and Application of Curves and Surfaces* —

- Mathematics of Surfaces II*, number 11 in The Institute of Mathematics and its Applications Series, pages 299–319, Cardiff, UK, 1987. Clarendon Press.
- [133] J. G. Semple and G. T. Kneebone. *Algebraic Projective Geometry*. Oxford Classic Texts in the Physical Sciences. Clarendon Press, Oxford, UK, 1998. Originally published in 1952.
- [134] D. F. Shanno. Conditioning of quasi-Newton methods for function minimization. *Maths. Comput.*, 24:647–656, 1970.
- [135] L. S. Shapiro, A. Zisserman, and M. Brady. 3D motion recovery via affine epipolar geometry. *Int. Journal of Computer Vision*, 16(2):147–182, October 1995.
- [136] A. Shashua. Trilinearity in visual recognition by alignment. In J.-O. Eklundh, editor, *Proc. 3rd European Conf. on Computer Vision*, volume I of *Lecture Notes in Computer Science 800*, pages 479–484, Stockholm, Sweden, May 1994. Springer-Verlag.
- [137] A. Shashua and M. Werman. Trilinearity of three perspective views and its associated tensor. In *Proc. 5th Int. Conf. on Computer Vision*, pages 920–925, Cambridge, USA, June 1995.
- [138] A. Shashua and L. Wolf. On the structure and properties of the quadfocal tensor. In D. Vernon, editor, *Proc. 6th European Conf. on Computer Vision*, volume I of *Lecture Notes in Computer Science 1842*, pages 710–724, Dublin, Ireland, June/July 2000. Springer-Verlag.
- [139] S. M. Smith and J. M. Brady. Susan: A new approach to low-level image-processing. *Int. Journal of Computer Vision*, 23(1):45–78, May 1997.

-
- [140] C. E. Springer. *Geometry and analysis of projective spaces*. Freeman, San Francisco, USA, 1964.
- [141] P. Sturm. Critical motion sequences for monocular self-calibration and uncalibrated Euclidean reconstruction. In *Proc. Conf. Computer Vision and Pattern Recognition*, pages 1100–1105, San Juan, Puerto Rico, June 1997.
- [142] P. Sturm. Critical motion sequences for the self-calibration of cameras and stereo systems with variable focal length. In T. Pridmore and D. Elliman, editors, *Proc. British Machine Vision Conference*, volume I, pages 63–72, Nottingham, September 1999.
- [143] P. Sturm and B. Triggs. A factorization based algorithm for multi-image projective structure and motion. In B. Buxton and R. Cipolla, editors, *Proc. 4th European Conf. on Computer Vision*, volume II of *Lecture Notes in Computer Science 1064*, pages 709–720, Cambridge, England, April 1996. Springer-Verlag.
- [144] M. Subbarao. Parallel depth recovery by changing camera parameters. In *Proc. 2nd Int. Conf. on Computer Vision*, pages 149–155, Tarpoon Springs, USA, December 1988.
- [145] R. Szeliski. Rapid octree construction from image sequences. *Computer Vision, Graphics and Image Processing*, 58(1):23–32, July 1993.
- [146] R. Szeliski and R. Weiss. Robust shape recovery from occluding contours using a linear smoother. *Int. Journal of Computer Vision*, 28(1):27–44, June 1998.

-
- [147] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization method. *Int. Journal of Computer Vision*, 9(2):137–154, November 1992.
- [148] P. H. S. Torr and A. Zisserman. Robust parametrization and computation of the trifocal tensor. *Image and Vision Computing*, 15(8):591–605, May 1997.
- [149] P. H. S. Torr and A. Zisserman. Feature based methods for structure and motion estimation. In B. Triggs, A. Zisserman, and R. Szeliski, editors, *Vision Algorithms: Theory and Practice*, number 1883 in Lecture Notes in Computer Science, pages 278–294, Corfu, Greece, September 1999. Springer-Verlag.
- [150] B. Triggs. Factorization methods for projective structure and motion. In *Proc. Conf. Computer Vision and Pattern Recognition*, pages 845–851, San Francisco, USA, 1996.
- [151] B. Triggs. Autocalibration and the absolute quadric. In *Proc. Conf. Computer Vision and Pattern Recognition*, pages 609–614, San Juan, Puerto Rico, June 1997.
- [152] B. Triggs. Plane + parallax, tensors and factorization. In D. Vernon, editor, *Proc. 6th European Conf. on Computer Vision*, volume I of *Lecture Notes in Computer Science 1842*, pages 522–538, Dublin, Ireland, June/July 2000. Springer-Verlag.
- [153] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon. Bundle adjustment — A modern synthesis. In B. Triggs, A. Zisserman, and R. Szeliski, editors, *Vision Algorithms: Theory and Practice*, number 1883 in Lecture

- Notes in Computer Science, pages 298–372, Corfu, Greece, September 1999. Springer-Verlag.
- [154] H. P. Trivedi. Can multiple views make up for lack of camera registration? *Image and Vision Computing*, 6(1):29–32, February 1988.
- [155] R. Y. Tsai. An efficient and accurate camera calibration technique for 3D machine vision. In *Proc. Conf. Computer Vision and Pattern Recognition*, pages 364–374, Miami, June 1986.
- [156] R. Y. Tsai and T. S. Huang. Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces. *IEEE Trans. Pattern Analysis and Machine Intell.*, 6(1):13–26, January 1984.
- [157] S. Ullman. *The interpretation of visual motion*. MIT Press, Cambridge, USA, 1979.
- [158] R. Vaillant and O. D. Faugeras. Using extremal boundaries for 3D object modelling. *IEEE Trans. Pattern Analysis and Machine Intell.*, 14(2):157–173, February 1992.
- [159] T. Vieville and D. Lingrand. Using specific displacements to analyze motion without calibration. *Int. Journal of Computer Vision*, 31(1):5–29, February 1999.
- [160] T. Wada, H. Ukida, and T. Matsuyama. Shape from shading with interreflections under a proximal light-source: Distortion-free copying of an unfolded book. *Int. Journal of Computer Vision*, 24(2):125–135, September 1997.
- [161] K.-Y. K. Wong, P. R. S. Mendonça, and R. Cipolla. Reconstruction and motion estimation from apparent contours under circular motion. In T. Pridmore

- and D. Elliman, editors, *Proc. British Machine Vision Conference*, volume 1, pages 83–92, Nottingham, UK, September 1999.
- [162] X. Yan, X. Dong-hui, P. Jia-xiong, and D. Ming-yue. The unique solution of projective invariants of six points from four uncalibrated images. *Pattern Recognition*, 30(3):513–517, March 1997.
- [163] C. Zeller and O. Faugeras. Camera self-calibration from video sequences: The Kruppa equations revisited. Technical Report RR-2793, INRIA, France, February 1996.
- [164] Z. Zhang. Determining the epipolar geometry and its uncertainty: A review. *Int. Journal of Computer Vision*, 27(2):161–195, March/April 1998.
- [165] A. Zisserman, D. Forsyth, J. L. Mundy, and C. A. Rothwell. Recognizing general curved objects efficiently. In J. L. Mundy and A. Zisserman, editors, *Geometric Invariance in Computer Vision*, Artificial Intelligence Series, chapter 11, pages 228–251. MIT Press, Cambridge, USA, 1992.
- [166] A. Zisserman, D. Liebowitz, and M. Armstrong. Resolving ambiguities in auto-calibration. *Phil. Trans. Royal Soc. London A*, 356:1193–1211, 1998.
- [167] A. Zisserman, J. L. Mundy, D. A. Forsyth, J. Liu, N. Pillow, C. Rothwell, and S. Utcke. Class-based grouping in perspective images. In *Proc. 5th Int. Conf. on Computer Vision*, pages 183–188, Cambridge, USA, June 1995.

Index

- affine
 - approximation, 5, 20, 43, 44, 57, 58
 - validity of, 5
- camera, 4, **20**, 27, 28, 38, 52
 - canonical, 38
 - optical centre, 20
- epipolar geometry, *see* epipolar geometry, affine
- fundamental matrix, *see* fundamental matrix, affine
- transformation, 16, 25, 38
- algebraic error, *see* fundamental matrix, computation of, **36**
- apparent contour, *see* profile
- Augustine, St., 55

- Barrow, H. G., 57
- bundle adjustment, 85, 90, 109, 112
 - projective, 90, 91
- camera
 - affine, *see* affine, camera
 - calibration, xiii, 3, 18, 56, 109
 - canonical, 22, 23, 46
 - extrinsic parameters, 18
 - motion, xiii, 34, 57
 - critical, *see* self-calibration, critical motions
 - planar, 44
 - rotation, 33, 63
 - table, 72
 - translation, 33, 97
 - multiview systems, *see* trifocal tensor *and* quadfocal tensor
- optical axis, 98–100
- optical centre, 19, 20, 28, 33, 56, 61, 62, 70
 - epipolar plane, *see* epipolar geometry, epipolar plane
 - epipole, *see* epipolar geometry, epipole
- optical ray, 19, 32
- projective, xx, 4, 5, **17–19**, 20, 25, 28, 30
 - general form, 18
 - geometric representation, 19
 - parameterisation, 18
- rotation, 67
- self-calibration, *see* self-calibration
- stereo, *see* epipolar geometry
- Cartesian coordinates, xix, **10**
 - conversion to homogeneous coordinates, 10, *see also* Cartesian coordinates
 - vectorial representation, 10, 12, 33
- circular motion, xiii, **55–85**
 - fundamental matrix, *see* fundamental matrix, parameterisation, circular motion
 - horizon, 56, 66, 74, 75, 80, 81
 - rotation axis, 61–63, 65, 68, 71
 - image of, 56, 61, 63, 64, 66, 68, 70, 80
 - turntable sequence, 56, 79, 80
- conic, **13**
 - absolute, 91

- equation, 13
- matrix of a , 13
- projection of a quadric, as a , 22
- contour generator, 42, 83
 - definition, **41**
 - intersection of, *see* frontier point
- corners, 35
 - effect of noise, 35
 - matching of, xiii, 38, 45, 58, 82
- correspondences
 - line, *see* lines, matching of
 - point, *see* corners, matching of
- critical motions, *see* self-calibration, critical motions
- derivative
 - finite differences, computation from, 80, 121
 - geometric error, of, 38, **122–124**
 - Hessian of geometric error, 38, **124–125**
 - Huang and Faugeras constraint, of, 106, **127–128**
 - complexity, 105
 - resultant of a polynomial and its, 23
- duality, principle of, **13**
- edges, 35
 - Canny detector, 75
- elimination theory, 23
- epipolar geometry, xiii, **27–45**
 - affine, 27, **38–41**
 - epipole, 37
 - epipolar constraint, 29, 98
 - epipolar line, **29**, 36, 41, 44, 98
 - computation of, 74
 - corresponding, 68
 - epipolar plane, **29**
 - epipolar tangency, *see* profile, epipolar tangency
 - epipole, **29**, 49
 - computation of, 56, 74, 75
 - horizon, *see* circular motion, horizon
 - orientation of, 43
 - symmetry, 68, 77
 - estimation, *see* fundamental matrix, computation of
 - fundamental matrix, *see* fundamental matrix
 - profiles, from, xiii, 1
 - smooth surfaces, 45
 - affine, 41, **43–45**, 57
- epipole, *see* epipolar geometry, epipole
- essential matrix, **33**
 - constraints on entries, *see* self-calibration, Huang and Faugeras constraints
 - decomposition, 34, 92, 104
 - degrees of freedom, 34
 - fundamental matrix, relation to, xiii, 93
- Euclidean
 - distance, *see* geometric error
 - geometry, 10
 - relation to projective geometry, 9
 - motion, 56
 - reconstruction, *see* reconstruction, Euclidean
 - transformation, 16, 25, *see also* projective, transformation, stratification of
- Faugeras, O. D., 91, *see also* self-calibration, Huang and Faugeras constraints
- frontier point, **41**, 42, 57
- fundamental matrix, xiii, **28–45**, 49
 - affine, 27, **38–41**
 - degrees of freedom, 40
 - parameterisation, 40, 41
 - computation of, **34–38**, 56

- analytical, **34–36**, 38
 - numerical, **36**
- constraints, 35, 67, 93
- degrees of freedom, **29**, 69
- epipolar constraint, *see* epipolar geometry, epipolar constraint
- multicamera tensors, and, 45
- parameterisation, 37, 59, **72**, 122
 - circular motion, 65, 76
 - plane + parallax, **30–33**, 70, 76
- geometric error, **36–38**, 43, 76
 - derivative, *see* derivative, geometric error, of
 - Hessian, *see* derivative, Hessian of geometric error
 - transfer of points and lines, in, 51
- Grassman-Cayley algebra, 38
- harmonic homology, **60–64**
 - computation of, 72, 73
 - bitangents, 72
 - degrees of freedom, 68
 - parameterisation, 63
 - particular cases, *see* symmetry, skewed and symmetry, affine plane induced homography, 70, 71, *see also* fundamental matrix, plane + parallax
- Hartley, R., vii, 91
- Hobbes, Thomas, xix
- Homogeneous coordinates, xix
- homogeneous coordinates, **10–15**
 - conics in, *see* conic, matrix of a equivalence class, as an, 11, 15
 - linear transformation of, *see* projective, transformation
 - lines in, 12
 - planes in, *see* planes, homogeneous coordinates, in
 - points in, *see* points, homogeneous coordinates, in
 - quadrics in, *see* quadric, matrix of a
 - vectorial representation, 10, 11
- homography, plane induced, *see* fundamental matrix, plane + parallax
- Huang, T., *see* self-calibration, Huang and Faugeras constraints
- implicit curves, **23–24**
 - degree, 23
 - projected from implicit surfaces, 23, *see also* surface, implicit and elimination theory
 - tensorial representation, 23
- implicit surfaces, *see* surface, implicit
- intrinsic parameters, xiii, 18, 56, 64, **92**
 - varying, xiii, 93
- Koch, R., 91
- Koenderink, J. J., 57
- Kruppa, E., *see* self-calibration, Kruppa equations
- level-set methods, 83
- lines
 - epipolar, *see* epipolar geometry, epipolar line
 - homogeneous coordinates, in, **12–13**
 - intersection of, 12
 - matching of, xiii, 45, 49
 - projection of planes, *see* planes, projection of
 - transfer of, *see* trifocal tensor, transfer of lines and points
- Luong, Q., 91
- matrix
 - essential, *see* essential matrix
 - fundamental, *see* fundamental matrix

- homogeneous, 15
- intrinsic parameters, *see* intrinsic parameters
- rotation, 17, *see also* Euclidean, transformation
- singular value decomposition (svd) of, 93
- Maybank, S., 91
- motion
 - critical, *see* self-calibration, critical motions
 - estimation from profiles, *see* profile, motion estimation from
 - structure and
 - direct methods, 2
 - feature based methods, 2
- optimisation
 - BFGS, 106
 - geometric error, minimisation of, *see* geometric error
 - Levenberg-Marquardt, 37, 106
 - Newton-Rhapon, 37
- planes
 - epipolar, *see* epipolar geometry, epipolar plane
 - homogeneous coordinates, in, **12–13**
 - points, defined by, 13
 - projection of, **20–21**
- Plato, 9
- points
 - Cartesian coordinates, in, 10
 - frontier, *see* frontier point
 - homogeneous coordinates, in, 10
 - invariant of six, 48
 - matching of, *see* corners, matching of
 - projection of, 18, 28, *see* camera, optical ray
 - transfer of, *see* trifocal tensor, transfer of lines and points
- Pollard, S. B., 57
- Pollefeys, M., 91
- Porrill, J., 57
- profile, xiii, **55–85**
 - definition, 41
 - epipolar tangency, **41**, 42, 43
 - feature, 3, 35
 - motion estimation from, 3, 42, 56, 57
 - affine, 43, 45, *see also* epipolar geometry, smooth surfaces, affine
 - local minima, 44
 - surface of revolution, of, *see* surface, revolution of, profile of *and* harmonic homology
- projective
 - camera, *see* camera
 - geometry, 4, **9–26**
 - definition, 9
 - relation to Euclidean geometry, 9
 - transformation, 3, **15–17**
 - 2-dimensional, 62
 - 3-dimensional, 91
 - invariance of fundamental matrix to, 30, *see also* fundamental matrix
 - quasi-affine, 39
 - stratification of, 16, 17
- quadfocal tensor, 27, 46
 - parameterisation, 53
 - quadrilinear constraint
 - relation to bilinear (epipolar) and trilinear constraint, 46
 - textbf, 52
- quadric
 - absolute, 90, 91
 - degenerate, **14**
 - matrix of a, 14
 - projection of a, 22

- tangents to a, 15
- reconstruction
- Euclidean, xiii, 91, *see also* bundle adjustment
 - profiles, from, 57, **82–85**
 - projective, 91, *see also* bundle adjustment, projective
 - quasi-affine, 39, *see also* projective, transformation, quasi-affine
- registration, image, 69
- Rieger, J. H., 57
- self-calibration, 1, 3, **89–115**
- critical motions, 4, **97**, 103, 110, 114
 - detection of, **97–100**, 114
 - Huang and Faugeras constraints, 89, **92–94**, 114
 - Kruppa equations, 90, 91
 - equivalence to Huang and Faugeras and Trivedi constraints, 90
 - minimal requirements for, 103
 - software for, 115
 - Trivedi constraints, 90
- similarity transformation, 16, *see also* projective, transformation, stratification of
- surface
- implicit, **23**
 - projection of, 23
 - tensorial representation, 23
 - revolution, of, 5, 6, 56, 58
 - profile of, 60, 64, 71, *see also* harmonic homology
 - smooth, 5
- Sylvester matrix, 23, *see also* derivative, resultant of a polynomial and its
- symmetry, 58, *see also* surface, revolution, of
- bilateral, 64
 - projective, *see* harmonic homology
 - skewed, 63, 79
- Tenenbaum, J. M., 57
- triangulation
- profiles, of, 58, 83, 85, *see* reconstruction, profiles, from
- trifocal tensor
- affine, 27, **51–52**, *see also* affine degrees of freedom, 52
 - parameterisation, 52
 - computation of, **49–51**
 - analytical, 49
 - numerical, 51
 - constraints, 49
 - degrees of freedom, 48
 - fundamental matrix, comparison to, 47
 - parameterisation, **47–49**
 - minimal, **48–49**
 - transfer of lines and points, 47, 51
 - trilinear constraint, **47**
- Triggs, W., 91
- Trivedi, H. P., *see* self-calibration, Trivedi constraints
- Van Gool, L., 91
- vanishing points, 66, 72
- voxel-carving, 83
- Zhang, Z., 28