Camera Calibration and Reconstruction of Geometry from Images

David Liebowitz
Merton College

Robotics Research Group
Department of Engineering Science
University of Oxford

Trinity Term 2001

David Liebowitz
Merton College

Doctor of Philosophy
Trinity Term 2001

Abstract

This thesis addresses the issues of combining camera calibration constraints from various sources and reconstructing scene geometry from single and multiple views. A geometric approach is taken, associating both structure recovery and calibration with geometric entities.

Three sources of calibration constraints are considered: scene constraints, such as the parallelism and orthogonality of lines, constraints from partial knowledge of camera parameters, and constraints derived from the motion between views.

First, methods of rectifying the projective distortion in an imaged plane are examined. Metric rectification constraints are developed by constraining the imaged plane circular points.

The internal camera parameters are associated with the absolute conic. It is shown how imaged plane circular points constrain the image of the absolute conic, and are constrained by a known absolute conic in return. A method of using planes with known metric structure as a calibration object is developed.

Next, calibration and reconstruction from single views is addressed. A well known configuration of the vanishing points of three orthogonal directions and knowledge that the camera has square pixels is expressed geometrically and subjected to degeneracy and error analysis. The square pixel constraint is shown to be geometrically equivalent to treating the image plane as a metric scene plane.

Use of the vanishing point configuration is extended to two views, where three vanishing points and known epipolar geometry define a three dimensional affine reconstruction. Calibration and metric reconstruction follows similarly to the single view case, with the addition of auto-calibration constraints from the motion between views. The auto-calibration constraints are derived from the geometric representation of the square pixel constraints, by transferring the image plane circular points between views. Degenerate cases for constraints from square pixels and cameras having identical internal parameters are described.

Finally, a constraint on the metric rectification of an affine reconstruction from the relative lengths of a pair of 3D line segments is developed. The constraint is applied to human motion capture from a pair of affine cameras.

Dedication

For my father, Frederick Liebowitz.

Acknowledgements

My years in Oxford have been an apprenticeship under a master of the craft, Andrew Zisserman. His insight and guidance have shaped this thesis, and I remain deeply grateful for all that he has taught me. Andrew Fitzgibbon has been a great source of help in all software matters and also patiently explained a great deal of vision to me. Most of all I would like to thank him for his boundless and infectious enthusiasm. Most of the last year was spent as a guest at the Royal Institute of Technology in Stockholm. I would like to thank Stefan Carlsson for making this a thoroughly rewarding and enjoyable experience.

I learned a great deal from my colleagues, and friends, in the Visual Geometry Group. Fred Schaffalitzky, Phil Pritchett, Antonio Criminisi, David Capel, Geoff Cross and Nic Pillow created a lively environment for research that I continue to miss. I would also like to thank Eric Hayman, Matthew Bryant and Sebastien Rougeaux for proofreading parts of this thesis.

Contents

Chapter 1

Introduction

A camera is an extraordinarily useful measuring device – it not only produces a realistic picture of a scene, but also provides information from which geometric properties of the scene can be measured. Reconstructing scene geometry from images is one of the most active areas in computer vision. It has a wide range of application in such diverse fields as architecture, archaeology, art history, and forensic science. In addition, the expansion of the Internet has fueled demand for greater realism in the information communicated online. Virtual reality environments and three dimensional product catalogues have become feasible and commercially attractive. Evolving standards such as Java, VRML and the humanoid animation standard HAnim support this trend by serving as a technological substrate for the explosion in the uses of vision technology.

Reconstruction techniques all require some form of calibration. This thesis addresses (mostly) the case where there is rich, regular scene geometry to provide calibration information. The novel contributions of the thesis will be addressed more precisely in the conclusion. In broad terms, however, the methods described integrate constraints on calibration and reconstruction of world geometry from different sources into a single framework. There are three sources of information considered:

1. scene constraints, such as the parallelism and orthogonality of lines and the known relative lengths of line segments,

2. internal parameter constraints, obtained when some of the internal camera parameters are

known, and

3. motion constraints, derived from the motion between views.

The objective is both to develop new calibration constraints and methods, and to do so in a manner which makes them easy to combine. Reconstruction methods from single and multiple views will be described and applied to a range of different scene and camera scenarios.

1.1 An overview

Visual perception of three dimensional information from a single image is a well understood phenomenon – photographs, even paintings, accurately convey depth information about the world on a two dimensional surface. As early as 1436 the renowned Alberti described a method of producing a painting with correct perspective effects: an artist views a scene monocularly through a piece of glass, draws on the glass and then transfers the drawing to another medium[77].[1] Shortly after the invention of *Alberti's window*, Leonardo da Vinci produced detailed instructions for artists expounding on perspective effects[60]. Leonardo went beyond the projection of geometry, including descriptions of the shading and colour effects an artist needs to take into account to realistically convey depth information. Koenderink[61] relates that, despite this knowledge, there was until relatively recently a complete denial in the scientific community that depth perception from monocular images was possible. It is now known that it is not only possible, but can also be very strong. Monocular depth perception is accomplished by the integration of numerous sources[77] such as shading, cast shadows, atmospheric perspective and geometric perspective effects.

Geometric perspective has received the most attention in the photogrammetry and vision literature, although shape from contour[8], texture[111] and shading[53] have been studied and there has even been preliminary work on depth from atmospheric effects[74]. Geometric perspective provides very powerful depth cues, particularly the perspective effects inherent in what the visual perception literature refers to as the *carpentered world*: a space dominated by parallel lines in orthogonal direc-

[1] In fact, the earliest perspectively correct painting is generally acknowedged to be Massacio's *la Trinita*, completed in 1427[59].

tions. In the famous Ames room illusion in figure 1.1, for example, other depth cues and distorted parallelism and orthogonality create a size illusion[55].

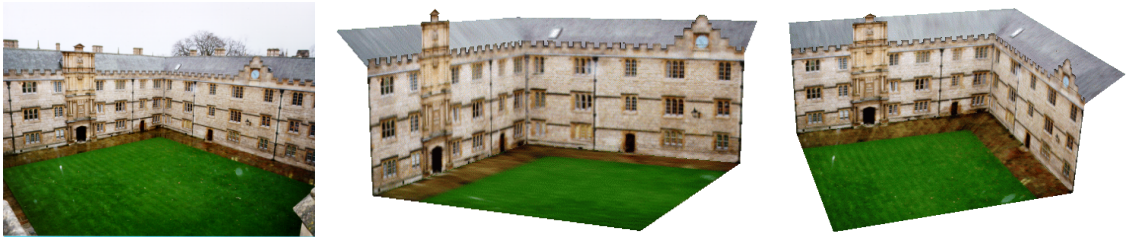
Figure 1.1: The Ames room illusion, where three men appear to have remarkably different sizes (from Palmer[77]). The Ames room appears to be a cuboid, but is in fact not. The left corner of the room, where the smallest figure appears, is further from the viewer than the right corner, and lower. The varying height of the floor of the room creates the illusion that all three men are standing the same distance from the viewer. The perception is reinforced by the features on the walls, which are distorted to appear to the viewer to be rectangular.

The parallelism and orthogonality inherent in the way humans construct the built environment means that there are plentiful constraints on calibration and reconstruction to be had from images of buildings or taken inside buildings. This information will be used here in a number of ways.

The first is correction of perspective warping in imaged planes. The distortion that plane geometry undergoes when photographed can be recovered up to an ambiguity depending on the available constraints. Such rectification has application to object recognition, image mosaicing and graphics, and also provides constraints in the calibration of the camera itself.

With a single view of three mutually orthogonal planes and some knowledge of the camera parameters, complete internal camera calibration can be achieved. This calibration together with the plane rectification techniques allow 3D reconstruction of objects such as buildings from a single view without scene measurements. The courtyard shown in figure 1.2, for example, can be approximated by a set of planes. Each of the planes may be rectified and relative orientations, where not

assumed orthogonal, computed. The result is a 3D model of the building from which novel views can be created.



(a) (b) (c)

Figure 1.2: 3D reconstruction from a single image. (a) Fellows quad, Merton College, Oxford. (b) (c) Views of the textured 3D model created from the single image in (a).

Such plane based single view reconstruction has useful applications. It may, for example, be applied to cases where it is not possible to take measurements of a scene or capture many views of an object in order to use stereo techniques. Reconstructing buildings from photographs to create models of valuable architectural sites is one such case, particularly where buildings have been destroyed and only archive images are available (see, for example, figure 4.26 in chapter 4).

Monocular reconstruction is not limited to planes, but can be accomplished for any geometry for which the parameters can be recovered from a single view. The Facade system of Debevec *et al.*[25] and the commercial package Canoma, for instance, allow inclusion of a number of geometric primitives in a reconstruction. There is, however, a substantial difference between single view reconstruction and what can be obtained from multiple views.

A single view reconstruction is restricted in the sense that knowledge of the 3D geometry is limited to the geometric primitives. In the case of scene planes, for example, the geometry on the planes and the relationships between planes can be computed. A complete 3D reconstruction of the scene, however, is only possible from two or more views, where in principle a dense reconstruction of an arbitrary shape can be computed. With multiple views the accuracy of reconstruction is improved and occlusion of parts of the scene can be dealt with. In addition, geometric hypotheses such as the co-planarity of a set of 3D points can be tested.

Model and stereo based methods can be combined. The model based single view calibration approach is accordingly extended to two views with the addition of constraints from the motion

between views. Such constraints take the form of the transfer of internal calibration information between views.

Scene structure constraints are also not restricted to built environments. Calibration and reconstruction information can be obtained from dynamic human motion.

Motion capture from image sequences is an established industry with numerous applications. The most prominent of these is in the entertainment sector, where motion capture from live actors is used to realistically animate graphics characters in films and games. Motion capture data is also used in clinical gait analysis and athletic performance evaluation. A detailed examination of movement aids rehabilitation of people with mobility impairment or helps athletes improve their techniques.

Motion capture systems are usually carefully calibrated to ensure accurate 3D measurement. An uncalibrated approach will be described here where the body itself is used to generate calibration constraints. The constraint mechanism exploits the fact that parts of the body such as the upper and lower arms and legs have a fixed length throughout the sequence. Figure 1.3 shows an example of a pair of frames from a sequence of an athlete doing a triplejump and a frame from the motion reconstruction obtained with this approach. Uncalibrated motion capture is particularly applicable to sport scenarios because of the difficulty of setting up systems of calibrated cameras.

(a) (b) (c)

Figure 1.3: (a) and (b) Two views from a pair of sequences of an athlete performing the triplejump. (c) A view from the reconstructed motion sequence.

All the constraints used, whether from structure, motion or known parameters, are described geometrically. The consequent relationships between constraints and ways of combining them are themes explored throughout the thesis.

1.2 Thinking in pictures

The approach taken in this thesis is determinedly geometric throughout, associating algebraic constraints with geometric entities that can be computed to effect reconstruction and calibration.

The specific use of geometry has the analytical advantage of linearizing many of the constraints that are used. Another motivation, and one that should not be underestimated, is that a geometric understanding of a problem provides an extremely useful way of thinking about it. Being able to visualize and draw pictures is a powerful aid to understanding. In their book on a statistical approach to geometric methods, Saville and Wood[88] state that when R.A. Fisher developed analysis of variance and regression methods, he did so geometrically and "derived his inventive inspiration from thinking in pictures". It has since become the norm to express these ideas only in algebraic form which, in the opinion of this author, makes them far more difficult to understand. Indeed, given a clear geometric representation of a problem, it should be much simpler to deal with (see figure 1.4).

Figure 1.4: With a sufficiently geometric understanding, it becomes child's play or: *the author's cunning plan to include a picture of his son is his thesis.*

Fortunately, the mathematics of calibration and reconstruction lends itself very well to geometric analysis and interpretation. The spatial configuration of cameras, models of projective spaces and an internal model of the camera itself have accessible geometric meaning. In particular, the model of camera internal parameters that is used for a pinhole camera can be associated with a very special conic: the absolute conic. Much of this thesis is concerned with methods of finding the absolute

conic.

1.3 Thesis structure

The overall layout of subsequent chapters is as follows. Chapter 2 briefly describes the necessary background material. The camera model and the relevant mathematics of projective spaces are introduced and the concepts of uncalibrated reconstruction presented. A broad survey of the calibration literature concludes the chapter. Specific reference to more closely related work is delayed until later chapters, where it can be described in context.

The theory of plane rectification from scene geometry is presented in chapter 3. Earlier work on recovering the affine geometry by computing the vanishing line of the plane is extended to metric geometry by obtaining constraints from the angles between lines in the plane and known relative lengths of line segments. Specific attention is paid to the computation of vanishing points from parallel lines since such computations occur frequently throughout the thesis. A maximum likelihood estimate of vanishing points is described and a first order error analysis performed.

Chapter 4 moves to computing the calibration of a camera from a single view using a combination of constraints. Sources of information considered include parallel lines in orthogonal directions, scene planes and partial prior knowledge of camera parameters. It is shown how calibration is used to rectify previously *partially* rectified planes and compute the relative orientation of planes. This information is then used to compute 3D reconstructions. A thorough degeneracy analysis of the particular case of viewing a cuboid object with a square pixel camera is performed. Error analysis of this configuration supports the degeneracy analysis. It is subsequently shown how knowledge of the principal point of the camera can be used in degenerate cases. The use of planes with known metric structure to calibrate a camera is described and an error analysis performed.

Chapter 5 applies the calibration techniques developed in chapter 4 to the two view case, where the *motion* between views provides further calibration information. A stratified reconstruction method for two views is presented and a measurement system described.

Motion capture using rigid links is described in chapter 6. A general constraint on calibration from the known ratio of lengths of a pair of line segments is derived. A motion capture application

reconstructing the movement of athletes from broadcast video footage using an affine camera model is then demonstrated.

Chapter 7 draws the thesis to a close, summing up the contributions made in the research and outlining areas of further investigation.

The work in this thesis is based on five papers published over the course of the research. Four of these[67, 117, 66, 68] make up the bulk of the material, and describe work done with Professor Andrew Zisserman. The fifth[65], the subject of chapter 6, was written in collaboration with Dr Stefan Carlsson while a guest at the Royal Institute of Technology in Stockholm.

Chapter 2

Background

2.1 Introduction

The material in this chapter introduces three background areas relevant to the thesis. The first section presents notation and results in the algebraic description of the projective geometry of the line, plane and three dimensional space. Particularly important are the metric and affine specializations of the 2D and 3D projective spaces and the geometric invariants with which they are associated. Properties of plane conics are enumerated in some detail, since they will be used frequently.

The second subject is the pinhole camera model and fundamental results in uncalibrated reconstruction from perspective cameras. This includes epipolar geometry and the stratified reconstruction scheme.

The third section is a brief overview of the camera calibration and reconstruction literature. The section is intended to remind the reader of the wide range of methods for camera calibration and auto-calibration, and thus to place the material in the following chapters in context. Reference to directly relevant work is delayed until later chapters.

Throughout the thesis, scalar quantities will be written in italics, such as s. Vector quantities are represented as column vectors indicated by a boldface Roman font, for example a 2D point $\mathbf{x}$ or 3D point $\mathbf{X}$. Matrices appear in a sans serif font, as in the perspective camera P.

2.2 Projective spaces and transformations

Projective geometry is an indispensable tool in the modeling and analysis of vision problems. Using projective notation and concepts has numerous advantages, among which is that points at infinity in a projective space may be dealt with in the same manner as other points, avoiding limiting procedures and special cases. Projective geometry also allows for compact linear algebra descriptions of the geometric entities and relationships which are encountered in vision problems.

A general n dimensional projective space is defined by $n+2$ basis points, and a point in the space is represented by an $n+1$ dimensional homogeneous vector defined up to an arbitrary non-zero scale factor. Projective transformations of the space are represented by $(n+1) \times (n+1)$ homogeneous matrices, similarly defined up to scale. Specific representations for geometric entities in one, two and three dimensions are described below. Note that, in equations relating homogeneous entities, '=' is used to indicate equality up to scale.

Most of the material in this section is taken from the now classic textbook of Semple and Kneebone[91]. The treatment is not meant to be rigorous, but to present the working tools that will be met in later chapters. The original text should be consulted for a thorough description of algebraic projective geometry.

2.2.1 The projective line

A point on the projective line, $\mathbb{P}^1$, is represented by a homogeneous two-vector $\mathbf{p} = (p_1,\, p_2)^\intercal$, with the scalar *projective parameter* defined to be $p = \frac{p_1}{p_2}$.

A homography of $\mathbb{P}^1$ is represented by 2×2 matrix H that transforms a point $\mathbf{p}$ as

$$\mathbf{p}' = \mathsf{H}\mathbf{p}$$

An important result defined on $\mathbb{P}^1$, and one that will be used in later chapters, is the invariance of the cross ratio. The cross ratio of two pairs of points $\mathbf{p}_a$, $\mathbf{p}_b$ and $\mathbf{p}_c$, $\mathbf{p}_d$ with projective parameters

a, b, c and d is defined as

$$\mathfrak{C}\{\mathbf{p}_a, \mathbf{p}_b; \mathbf{p}_c, \mathbf{p}_d\} = \frac{a-c}{a-d} : \frac{b-c}{b-d}$$

The cross ratio is invariant to projective transformations of $\mathbb{P}^1$: given homography H and points $\mathbf{p}'_a = \mathsf{H}\mathbf{p}_a$ and similarly for $\mathbf{p}'_b$, $\mathbf{p}'_c$ and $\mathbf{p}'_d$, with projective parameters a', b', c' and d', the cross ratio remains constant:

$$\frac{a-c}{a-d} : \frac{b-c}{b-d} = \frac{a'-c'}{a'-d'} : \frac{b'-c'}{b'-d'}$$

The pairs of points $\mathbf{p}_a$, $\mathbf{p}_b$ and $\mathbf{p}_c$, $\mathbf{p}_d$ are called *harmonic* if

$$\mathfrak{C}\{\mathbf{p}_a, \mathbf{p}_b; \mathbf{p}_c, \mathbf{p}_d\} = -1$$

The harmonic relation is associated with the orthogonality of directions in higher dimensional spaces.

Note that it is permissible, with the cross ratio defined as above, to write the projective parameter of point $(1,\ 0)^\intercal$ as ∞, and to use this in cross ratio computations.

2.2.2 The projective plane

Points and lines

A point on the projective plane $\mathbb{P}^2$ is represented by a homogeneous three-vector: $\mathbf{x} = (x_1,\ x_2,\ x_3)^\intercal$. The inhomogeneous finite point $\tilde{\mathbf{x}}$ is found from the elimination of the scale factor and is given by the two-vector

$$\tilde{\mathbf{x}} = (\tilde{x}_1, \tilde{x}_2)^\intercal = (\frac{x_1}{x_3}, \frac{x_2}{x_3})^\intercal$$

A line on the plane is defined by a three-vector such that, for line $\mathbf{l} = (l_1,\ l_2,\ l_3)^\mathsf{T}$ and point on the line $\mathbf{x}$

$$\mathbf{x}^\mathsf{T}\mathbf{l} = 0$$

The point of intersection of two lines $\mathbf{l}_1$ and $\mathbf{l}_2$ is given by

$$\mathbf{x} = \mathbf{l}_1 \times \mathbf{l}_2$$

and the line passing through two points $\mathbf{x}_1$ and $\mathbf{x}_2$ is

$$\mathbf{l} = \mathbf{x}_1 \times \mathbf{x}_2$$

These equations are an example of the duality that exists for points and lines in the plane: any theorem of $\mathbb{P}^2$ has a dual result obtained by switching the roles of points and lines.

The cross ratio is defined for four co-linear points on the plane, and dually for four concurrent lines on the plane. The cross ratio of a pencil of four lines can be viewed geometrically (see figure 2.1) as the cross ratio of the four points on any line intersecting all four lines of the pencil (except at the point of concurrency). The cross ratio is invariant to the choice of intersecting line.

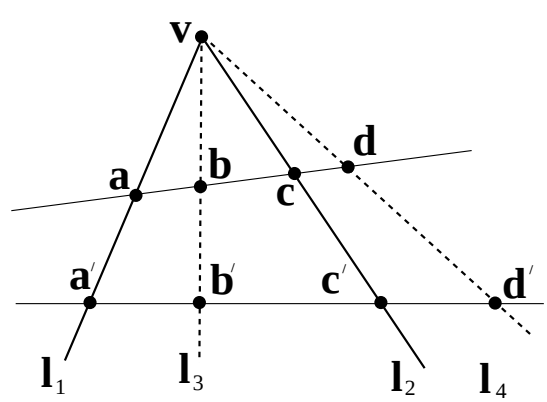


Figure 2.1: The cross ratio of four concurrent lines. Four lines $\mathbf{l}_1$ to $\mathbf{l}_4$ intersect in point $\mathbf{v}$. The cross ratio of the four lines is the cross ratio of the four points of intersection of $\mathbf{l}_1$ to $\mathbf{l}_4$ with any line not through $\mathbf{v}$. That is $\mathfrak{C}\{\mathbf{l}_1, \mathbf{l}_2; \mathbf{l}_3, \mathbf{l}_4\} = \mathfrak{C}\{\mathbf{a}, \mathbf{b}; \mathbf{c}, \mathbf{d}\} = \mathfrak{C}\{\mathbf{a}', \mathbf{b}'; \mathbf{c}', \mathbf{d}'\}$.

A projective transformation of the plane is represented by a 3×3 homogeneous matrix which transforms a point $\mathbf{x}$ to

$$\mathbf{x}' = \mathsf{H}\mathbf{x}$$

The corresponding transformation of a line $\mathbf{l}$ is given by

$$\mathbf{l}' = \mathsf{H}^{-\mathsf{T}}\mathbf{l}$$

Conics

The curves described by conic sections feature prominently in the discussions that follow. Such curves can be written

$$C_1x_1^2 + C_2x_1x_2 + C_3x_2^2 + C_4x_1x_3 + C_5x_2x_3 + C_6x_3^2 = 0$$

More conveniently, a conic can be represented by a 3×3 symmetric matrix with the form[73]

$$\mathsf{C} = \begin{pmatrix} C_1 & C_2/2 & C_4/2 \\ C_2/2 & C_3 & C_5/2 \\ C_4/2 & C_5/2 & C_6 \end{pmatrix}$$

for which all points on the conic satisfy

$$\mathbf{x}^{\mathsf{T}}\mathsf{C}\mathbf{x} = 0$$

Under a point transformation $\mathbf{x}' = \mathsf{H}\mathbf{x}$, a conic C transforms as

$$\mathsf{C}' = \mathsf{H}^{-\mathsf{T}}\mathsf{C}\mathsf{H}^{-1} \tag{2.1}$$

A conic is the locus of the set of points lying on the conic curve. A dual representation is to consider the envelope of a set of tangent lines to the conic. All lines tangent to C satisfy

$$\mathbf{l}^{\mathsf{T}}\mathsf{C}^{-1}\mathbf{l} = 0$$

and C^{-1} is referred to as the dual of C.

Degenerate conics

Conic matrices describing circles, ellipses, hyperbolae and parabolas are rank three. It is, however, also possible to form rank deficient conic matrices to describe lines, and rank deficient dual conics to represent points.

A line $\mathbf{l}_1$ can be considered as a rank 1 degenerate conic. Writing the conic $\mathsf{L}_1 = \mathbf{l}_1\mathbf{l}_1^{\mathsf{T}}$, all points on $\mathbf{l}_1$ satisfy

$$\mathbf{x}^{\mathsf{T}}\mathbf{l}_1\mathbf{l}_1^{\mathsf{T}}\mathbf{x} = \mathbf{x}^{\mathsf{T}}\mathsf{L}_1\mathbf{x} = 0$$

A pair of lines define a rank two degenerate conic, written $\mathsf{L}_{12} = \mathbf{l}_1\mathbf{l}_2^{\mathsf{T}} + \mathbf{l}_2\mathbf{l}_1^{\mathsf{T}}$. All points on *either* line satisfy

$$\mathbf{x}^{\mathsf{T}}(\mathbf{l}_1\mathbf{l}_2^{\mathsf{T}} + \mathbf{l}_2\mathbf{l}_1^{\mathsf{T}})\mathbf{x} = \mathbf{x}^{\mathsf{T}}\mathsf{L}_{12}\mathbf{x} = 0$$

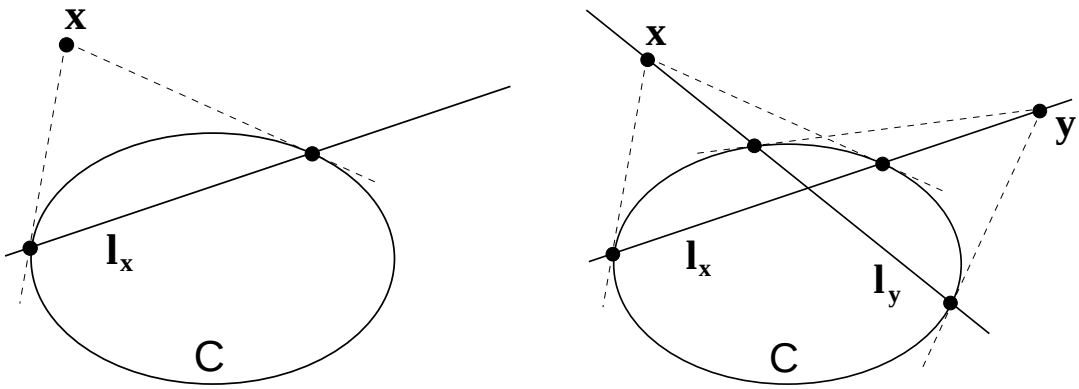
Similarly, a dual conic is defined for a single point $\mathbf{x}_1$. The pencil of all lines incident with $\mathbf{x}_1$ form the envelope of the dual conic $\mathsf{X}_1 = \mathbf{x}_1\mathbf{x}_1^{\mathsf{T}}$. For a pair of points $\mathbf{x}_1$ and $\mathbf{x}_2$, the dual conic $\mathsf{X}_{12} = \mathbf{x}_1\mathbf{x}_2^{\mathsf{T}} + \mathbf{x}_2\mathbf{x}_1^{\mathsf{T}}$ has as envelope the pencils of lines through each of the points.

Polarity and conjugacy

The polar of a point $\mathbf{x}$ with respect to a conic C is the line $\mathbf{l}_{\mathbf{x}} = \mathsf{C}\mathbf{x}$. Geometrically, the polar is the line through the points of tangency to C of a pair of lines through $\mathbf{x}$, as in figure 2.2 (a). Dually, $\mathbf{x} = \mathsf{C}^{-1}\mathbf{l}_{\mathbf{x}}$ is the pole of $\mathbf{l}_{\mathbf{x}}$ with respect to C.

A pair of points $\mathbf{x}$ and $\mathbf{y}$ are conjugate with respect to a conic C if $\mathbf{x}$ lies on the polar $\mathbf{l_y}$ of $\mathbf{y}$ with respect to C and $\mathbf{y}$ lies on the polar $\mathbf{l_x}$ of $\mathbf{x}$ with respect to C (figure 2.2 (b)). The conjugate points and lines satisfy

$$\mathbf{x}^\mathsf{T}\mathsf{C}\mathbf{y} = 0 \text{ and } \mathbf{l}_\mathbf{x}^\mathsf{T}\mathsf{C}^{-1}\mathbf{l_y} = 0$$



(a) (b)

Figure 2.2: (a) The polar of $\mathbf{x}$ with respect to C is $\mathbf{l_x}$. (b) Points $\mathbf{x}$ and $\mathbf{y}$ are conjugate with respect to conic C.

To compute the points of intersection of a line and a conic the Joachimsthal equation is used. Suppose the intersection of $\mathbf{l_y}$ with C is required. Any point on $\mathbf{l_y}$ can be parametrized as the sum of two points on the line, say $\mathbf{x}_1$ and $\mathbf{x}_2$. Points on the line $\mathbf{x}_1 + \lambda\mathbf{x}_2$ which also lie on $\mathbf{C}$ must satisfy

$$(\mathbf{x}_1 + \lambda\mathbf{x}_2)^\mathsf{T}\mathsf{C}(\mathbf{x}_1 + \lambda\mathbf{x}_2) = 0$$

The resulting quadratic in λ has two solutions. If C is a real conic, two distinct real solutions describe the two points of intersection, a single real solution indicates that the line is tangent to the conic and complex solutions indicate that the line and conic do not intersect.

Polarity and conjugacy are also defined for degenerate conics. For example, the polar of a point $\mathbf{x}$ with respect to the rank two degenerate conic L_{12} is the line $\mathbf{l_x} = \mathsf{L}_{12}\mathbf{x}$ through the point of intersection of $\mathbf{v} = \mathbf{l}_1 \times \mathbf{l}_2$ such that the two pairs of lines $\mathbf{l}_1$, $\mathbf{l}_2$ and $\mathbf{l_x}$, $(\mathbf{x} \times \mathbf{v})$ are harmonic. This implies that points $\mathbf{x}$ and $\mathbf{y}$ conjugate with respect to L_{12} are harmonic with respect to the points of

intersection of l_1 and l_2 with line $\mathbf{x} \times \mathbf{y}$ (see figure 2.3).

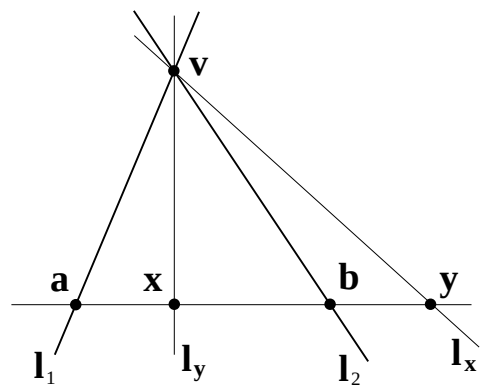


Figure 2.3: The polar of $\mathbf{x}$ with respect to a degenerate conic L_{12} is $\mathrm{l}_{\mathbf{x}}$, and similarly for $\mathbf{y}$. With $\mathbf{x}$ and $\mathbf{y}$ conjugate with respect to L_{12}, the point pairs $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{a}$, $\mathbf{b}$ are harmonic.

Specializations of the plane

Affine and metric properties of the projective plane are defined by specializing a line and a pair of points respectively. The affine specialization of the plane is obtained when a line is selected as the line at infinity. Affine properties are then defined. For example, set of concurrent lines are defined to be parallel if they intersect on the line at infinity.

The metric specialization of the plane is obtained from the affine plane by selecting a pair of complex conjugate points on l_∞. These points are known as the circular points[1] and are denoted $\mathbf{I}$ and $\mathbf{J}$. Once the circular points are selected, metric properties of the plane such as angle and relative length are defined.

The angle between a pair of lines, for example, is defined by their relationship to $\mathbf{I}$ and $\mathbf{J}$. Given a pair of lines l_a and l_b intersecting at $\mathbf{x}$, define the lines $\mathrm{l}_{\mathbf{I}} = \mathbf{x} \times \mathbf{I}$ and $\mathrm{l}_{\mathbf{J}} = \mathbf{x} \times \mathbf{J}$. The angle θ between l_a and l_b is given by Laguerre's formula: [31] as

$$\cos(2\theta) + i\sin(2\theta) = \mathfrak{C}\{\mathrm{l}_a, \mathrm{l}_b; \mathrm{l}_{\mathbf{I}}, \mathrm{l}_{\mathbf{J}}\} \tag{2.2}$$

[1]The circular points derive their name from the property that they are the intersection of every real circle in the plane with the line at infinity.

Choosing to observe the cross ratio of the two line pairs as the cross ratio of the four points of intersection of these lines with $\mathbf{l}_\infty$, (2.2) becomes

$$\cos(2\theta) + i\sin(2\theta) = \mathfrak{C}\{\mathbf{a}, \mathbf{b}; \mathbf{I}, \mathbf{J}\} \tag{2.3}$$

where $\mathbf{a}$ and $\mathbf{b}$ are the intersections of $\mathbf{l}_a$ and $\mathbf{l}_b$ with $\mathbf{l}_\infty$. If $\mathbf{l}_a$ and $\mathbf{l}_b$ are orthogonal, (2.3) becomes

$$\mathfrak{C}\{\mathbf{a}, \mathbf{b}; \mathbf{I}, \mathbf{J}\} = -1$$

and the pairs $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{I}$, $\mathbf{J}$ are harmonic. The identification of orthogonality with the harmonic cross ratio is useful in chapter 4.

By choosing the co-ordinates for the line at infinity and circular points to be $\mathbf{l}_\infty = (0,\ 0,\ 1)^\mathsf{T}$ and $\mathbf{I} = (1,\ i,\ 0)^\mathsf{T}$, $\mathbf{J} = (1,\ -i,\ 0)^\mathsf{T}$, the affine and metric specializations of the plane correspond to the familiar algebra of Euclidean geometry. Note that the canonical line at infinity is invariant to affine transformations of the plane, and the canonical circular points are invariant to similarity transformations of the plane.

Specializations of the plane in the context of image rectification is the subject of chapter 3, where it will also be shown how the line at infinity and the circular points can be considered and manipulated as a single entity.

2.2.3 The three dimensional projective space

The final projective space that is significant here is $\mathbb{P}^3$, the space of three dimensions. A point in $\mathbb{P}^3$ is defined by a homogeneous four-vector

$$\mathbf{X} = (X_1,\ X_2,\ X_3,\ X_4)^\mathsf{T}$$

with associated inhomogeneous point $\tilde{\mathbf{X}} = (\frac{X_1}{X_4},\ \frac{X_2}{X_4},\ \frac{X_3}{X_4})$. A plane is described by

$$\boldsymbol{\pi} = (\pi_1,\ \pi_2,\ \pi_3,\ \pi_4)^\mathsf{T}$$

with points on plane $\boldsymbol{\pi}$ satisfying

$$\boldsymbol{\pi}^\mathsf{T}\mathbf{X} = 0$$

The principle of duality applies to points and planes in $\mathbb{P}^3$.

Three points define a plane and three planes a point. Observing that

$$\begin{pmatrix} \mathbf{X}_1^\mathsf{T} \\ \mathbf{X}_2^\mathsf{T} \\ \mathbf{X}_3^\mathsf{T} \end{pmatrix} \boldsymbol{\pi} = \mathsf{A}_\mathbf{X}\boldsymbol{\pi} = \mathbf{0}$$

$\boldsymbol{\pi}$ can be computed as the null vector of $\mathsf{A}_\mathbf{X}$ or in terms of the cross products of inhomogeneous forms of $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{X}_3$[49]. A point can be computed from the intersection of three planes in exactly the same way.

Two points or two planes define a line. Lines occupy a notationally less convenient place in 3D, and can be represented in several ways: as the intersection of planes, as the span of co-linear points, by an antisymmetric matrix or using the Plücker co-ordinates[49]. Computations using lines are not considered further here.

A homography in 3D is represented by 4×4 homogeneous matrix which transform points as

$$\mathbf{X}' = \mathsf{H}\mathbf{X}$$

and planes as

$$\boldsymbol{\pi}' = \mathsf{H}^{-\mathsf{T}}\boldsymbol{\pi}$$

Quadrics

Quadric surfaces in 3D, such as spheres, paraboloids and cones, are largely analogous to conics in 2D. A quadric is represented by a 4×4 symmetric homogeneous matrix Q such that points on the

quadric satisfy

$$\mathbf{X}^{\mathsf{T}}\mathsf{Q}\mathbf{X} = 0$$

The dual representation of a quadric is defined by tangent planes satisfying

$$\boldsymbol{\pi}^{\mathsf{T}}\mathsf{Q}^{-1}\boldsymbol{\pi} = 0$$

As with conics, degenerate quadrics, with rank less than four, can represent a plane, pair of planes or a cone. Similarly, a point, point pair and *disk quadric* can be written as degenerate dual quadrics. Disk quadrics have particular relevance to calibration since a particular disk quadric encodes the specializations of $\mathbb{P}^3$.

A disk quadric is a rank three degenerate dual quadric. The plane $\boldsymbol{\pi}_{\mathsf{Q}}$ which is the null vector of the quadric matrix is in its envelope. The remaining planes in the envelope form a system of pencils of planes with axes on $\boldsymbol{\pi}_{\mathsf{Q}}$, and the lines of intersection of all these planes with $\boldsymbol{\pi}_{\mathsf{Q}}$ are tangent to a conic on $\boldsymbol{\pi}_{\mathsf{Q}}$ (see figure 2.4). It is helpful to visualize a disk quadric by considering the tangent planes to an ellipsoid that collapses into one of its planes of symmetry.

Polarity and conjugacy of quadrics will not be encountered explicitly in this thesis. For the sake of completeness, however, note that a point has a polar plane with respect to a quadric, while the polar of a line is a line.

Specializations of $\mathbb{P}^3$

Analogously to the plane, specializations of $\mathbb{P}^3$ are defined when particular geometric entities are distinguished. The affine space is defined when a plane is selected as the plane at infinity, and affine properties of the space are subsequently measurable. The metric specialization of 3D space occurs when the absolute conic on the plane at infinity is selected.

The absolute conic was introduced to vision by Faugeras[34], and plays an important conceptual role in camera calibration and scene reconstruction. The use of the absolute conic will be discussed in the following section, so it is worthwhile examining it more closely here.

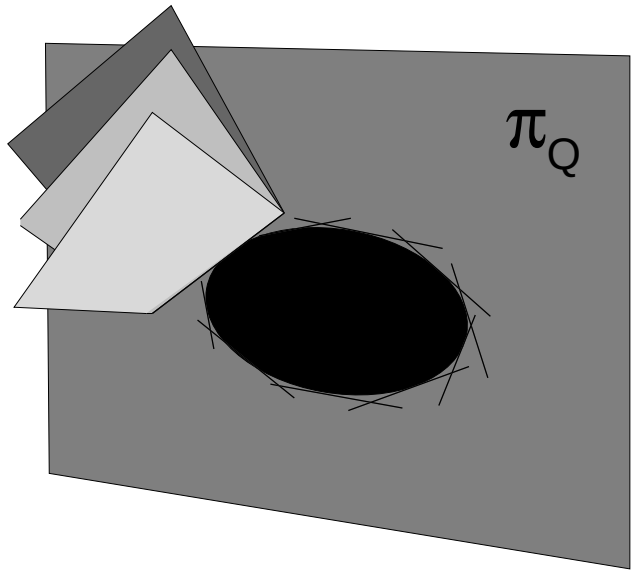


Figure 2.4: A disk quadric.

The canonical absolute conic Ω_∞ is often expressed by the equation

$$X_1^2 + X_2^2 + X_3^2 + X_4^2 = 0, \quad X_4 = 0$$

It can thus be considered as a circle on the canonical plane at infinity $\boldsymbol{\pi}_\infty = (0,\ 0,\ 0,\ 1)^\intercal$, albeit with radius i[31].

The corresponding symmetric matrix representation is

$$\Omega_\infty = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \mathsf{I}$$

Every plane in space intersects $\boldsymbol{\pi}_\infty$ in a line, and each of these lines intersects Ω_∞ in the circular points of the plane. This is illustrated in figure 4.2, chapter 4, where the property induces calibration constraints on a camera.

As with the line at infinity and circular points in the plane, Ω_∞ is invariant to similarity transformations of 3D and the plane at infinity is invariant to 3D affine transformations.

The absolute conic and the plane at infinity are encoded in a single entity, the absolute disk quadric Q_∞ [105] (also known as the absolute dual quadric). This disk quadric has the plane at infinity in its envelope, and the absolute conic as the conic formed by the tangent lines defined by the other planes in the envelope. The canonical disk quadric matrix representation is

$$\mathsf{Q}_\infty = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

2.3 Cameras and reconstruction

Knowing the internal and external camera parameters associated with a sequence of two or more images of a scene, it is possible to reconstruct geometry by back-projecting matched points in the images to three dimensional points in the world co-ordinate system. In fact, with only the internal parameters of each camera, metric reconstruction is possible [96]. In the absence of this or equivalent knowledge, however, reconstruction of geometry is only possible up to projective ambiguity. The notion of reconstruction without calibration was originally developed in the context of orthographic projection by Koenderink and van Doorn[62]. Their paper shows that uncalibrated views of a scene captured with affine cameras permit an affine reconstruction. This idea was extended to projective cameras by Faugeras[29] and Hartley *et. al.*[47], where a minimum of seven matched points in two views determine a reconstruction up to a projective transformation of the true world points. Calibration or auto-calibration of the cameras is the task of computing this projective transformation to rectify the reconstruction.

This section describes the pinhole model based camera, its use in projective reconstruction from image correspondences without calibration and an approach to subsequent metric reconstruction.

2.3.1 The pinhole camera

The pinhole camera models the process of projecting a 3D world point $\mathbf{X}$ to an image point $\mathbf{x}$ and is represented by a 3×4 projection matrix P:

$$\begin{aligned} \mathbf{x} &= \mathsf{P}\mathbf{X} \\ &= \mathsf{K}\,(\mathsf{R} \mid \mathbf{t})\,\mathbf{X} \end{aligned}$$

Rotation R and translation $\mathbf{t}$ describe the position and orientation of the camera relative to the world co-ordinate system. Matrix K encodes the *internal parameters* of the camera; the focal length f, skew k, aspect ratio r and principal point $(u_0,\ v_0)^\intercal$:

$$\mathsf{K} = \begin{pmatrix} f & k & u_0 \\ 0 & rf & v_0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2.4}$$

The principal point is the intersection of the optic axis of the camera with the image plane. Skew k is a factor dependent on the physical angle θ between the vertical and horizontal axes in the sensor array, given by $k = f\cot(\theta)$, and is often considered to be zero. The aspect ratio r refers to the relative vertical and horizontal scaling in the image. Although often assumed to be one, it may take on a different value as a result of the physical construction of the camera or digitisation hardware. A camera with zero skew and unit aspect ratio will be referred to as a *square pixel camera*.

The projective camera model describes a pinhole camera with the addition of internal parameters to account for real camera characteristics and an image based co-ordinate system. Physical lenses, however, introduce distortion in the image. This distortion is most often modeled by radial distortion[107], and sometimes by an additional tangential distortion component [99]. Distortion will be ignored in this thesis – it is insignificant in the examples that will appear, but can be corrected where necessary[27].

2.3.2 Epipolar geometry and reconstruction from uncalibrated cameras

Epipolar geometry describes the essential relationship between two cameras and their images. Consider the image of the optic centres of each of two cameras seen by the other, referring to figure 2.5. The optic centre of the first camera, $\mathbf{O}$, is imaged by the second camera in a point known as the epipole, $\mathbf{e}'$. The first view in turn images the optic centre of the second camera $\mathbf{O}'$ as epipole $\mathbf{e}$. The line segment joining $\mathbf{O}$ and $\mathbf{O}'$, intersecting the image planes in $\mathbf{e}$ and $\mathbf{e}'$, is the baseline between cameras. Now, world point $\mathbf{X}$ together with $\mathbf{O}$ and $\mathbf{O}'$ define a plane $\boldsymbol{\pi}_\mathbf{e}$. The intersection of $\boldsymbol{\pi}_e$ with the image planes defines two lines, $\mathbf{l}_\mathbf{e}$ and $\mathbf{l}'_\mathbf{e}$ in the two images, the epipolar lines. The images of $\mathbf{X}$, $\mathbf{x}$ and $\mathbf{x}'$, lie on the epipolar lines. These relations are summarized in the fundamental matrix F [29], a mapping between a point in one view and its corresponding epipolar line in the other view. Specifically,

$$\mathbf{l}'_\mathbf{e} = \mathsf{F}\mathbf{x} \quad \text{and} \quad \mathbf{l}_\mathbf{e} = \mathsf{F}^\top\mathbf{x}'$$

whence

$$\mathbf{l}'^\top_\mathbf{e}\mathbf{x} = \mathbf{x}'^\top\mathsf{F}\mathbf{x} = 0$$

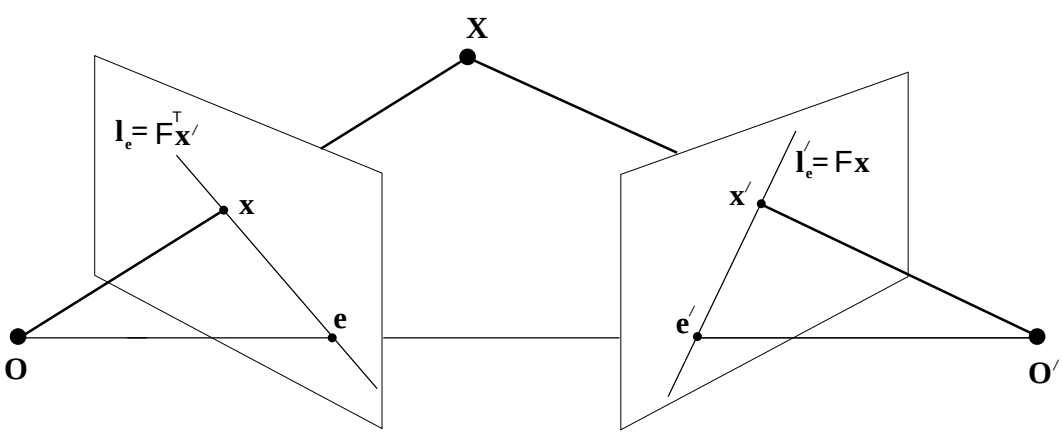


Figure 2.5: Epipolar geometry.

Epipolar geometry, methods of computing F and its implications have been extensively studied in vision. F can be computed non-linearly from seven points, linearly from eight, and a least-squares solution found from more than eight[49]. Figure 2.6 shows the pair of images of a building with three matched points and their epipolar lines. (The fundamental matrix for this view pair is obtained by a process described in section 5.3.)

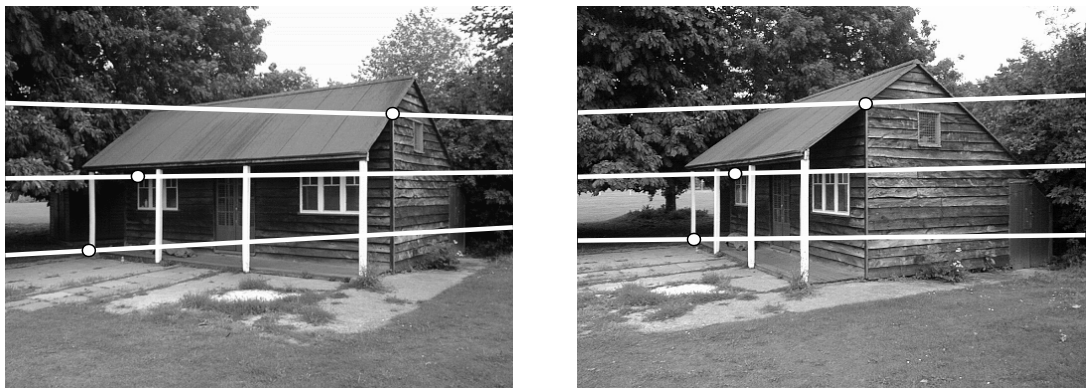


Figure 2.6: Two views of a building showing three corresponding points and their epipolar lines.

A reconstruction of the matched points follows from F. A pair of uncalibrated projection matrices, P_p and P'_p, are defined with P_p located at the origin of the reconstruction co-ordinate system:

$$\mathsf{P}_\mathsf{p} = (\mathsf{I} \mid \mathbf{0}) \tag{2.5}$$

It has been shown that the second projection can be written [69]

$$\mathsf{P}'_\mathsf{p} = (\mathsf{G} \mid \mathbf{e}') \tag{2.6}$$

where G is a 3×3 homography given by

$$\mathsf{G} = [\mathbf{e}']_\times \mathsf{F} - \mathbf{e}'\tilde{\boldsymbol{\pi}}_a^\mathsf{T}$$

Vector $\boldsymbol{\pi}_a = (\tilde{\boldsymbol{\pi}}_a^\mathsf{T},\ 1)^\mathsf{T}$ is a reference plane, the choice of which provides different projective reconstructions. An example projective reconstruction appears in figure 2.7.

A projective reconstruction is a representation of the geometry of the scene in $\mathbb{P}^3$. The geomet-

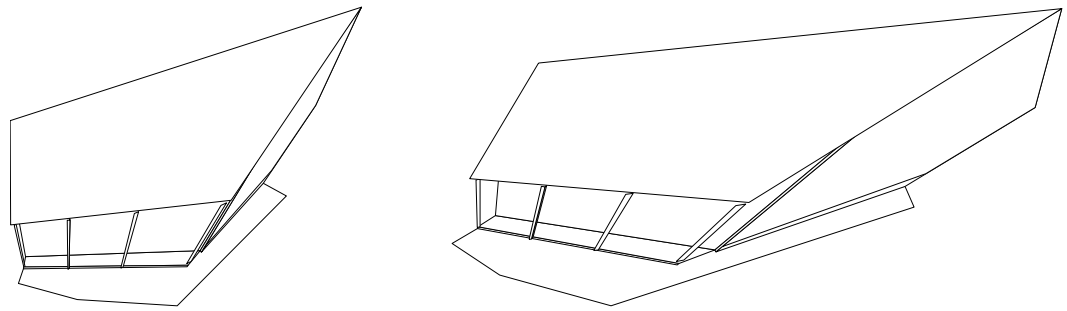


Figure 2.7: Two views of a projective reconstruction of the building of figure 2.6. Note that affine and metric properties of the structure are distorted – parallel lines in the world do not appear parallel and relative lengths and angles are incorrect.

ric entities, such as points, lines and planes, are assigned co-ordinates by the choice of $\tilde{\boldsymbol{\pi}}_a$, and are related to world structure by a general projective transformation H such that, for world point $\mathbf{X}_\mathsf{w}$ and projective reconstruction point $\mathbf{X}_\mathsf{p}$

$$\mathbf{X}_\mathsf{w} = \mathsf{H}\mathbf{X}_\mathsf{p} \tag{2.7}$$

The goal of a reconstruction algorithm is usually to obtain at least a metric reconstruction, one related to the scene by a similarity transformation. To accomplish this requires that the internal calibration of the cameras and their relative pose be computed. The following sections describe a stratified approach to metric reconstruction by way of an affine reconstruction.

2.3.3 Stratified metric rectification

The homography relating a projective reconstruction to world geometry can be decomposed into similarity, affine and projective transformations [32]. Formally, the rectifying homography H, which has 15 degrees of freedom, can be decomposed initially into metric and non-metric transformations:

$$\mathsf{H} = \mathsf{MN} \tag{2.8}$$

The metric part M is a similarity transformation of 3D of the form

$$\mathsf{M} = \begin{pmatrix} s\mathsf{R} & \mathbf{t} \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix}$$

where s is an isotropic scaling, R a rotation of 3D with three degrees of freedom and $\mathbf{t}$ a translation with three degrees of freedom. M thus has seven degrees of freedom. The non-metric component of H encodes the remaining eight degrees of freedom, and can in turn be decomposed in pure projective and affine transformations $\mathsf{H_p}$ and $\mathsf{H_a}$ respectively:

$$\mathsf{N} = \mathsf{H_a}\mathsf{H_p} \tag{2.9}$$

To obtain a metric reconstruction, only N need be computed. The geometric nature of the eight degrees of freedom in N will now be described in terms of specializations of $\mathbb{P}^3$.

2.3.4 Specialization in reconstruction

Given a particular projective reconstruction of a scene, an affine reconstruction can be obtained by specializing the projective space. Affine properties of the scene can then be measured if the choice of plane at infinity $\boldsymbol{\pi}_\infty$ is the true plane at infinity in the world. Similarly, metric scene properties can be obtained after selection of the true absolute conic.

Transformation of the structure and cameras by N to transform the plane at infinity and the absolute conic to their canonical representations results in a metric reconstruction. The components of N in (2.9), $\mathsf{H_p}$ and $\mathsf{H_a}$, are responsible for the plane at infinity and absolute conic respectively. The first step is to define $\mathsf{H_p}$ as the transformation that maps $\boldsymbol{\pi}_\infty$ to its affine canonical value $(0,\ 0,\ 0,\ 1)^\mathsf{T}$. It is simple to verify that this is accomplished by

$$\mathsf{H_p} = \begin{pmatrix} \mathsf{I} & \mathbf{0} \\ \boldsymbol{\pi}_\infty^\mathsf{T} & \end{pmatrix} \tag{2.10}$$

Following transformation by $\mathsf{H_p}$, the absolute conic Ω_∞ must be transformed to the 3×3 identity

matrix I. Noting that the upper left 3×3 matrix of a 4×4 homography of 3D is a plane homography of the canonical plane at infinity, $\mathsf{H_a}$ is the affine transformation of 3D

$$\mathsf{H_a} = \begin{pmatrix} \mathsf{U} & \mathbf{0}^\mathsf{T} \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix} \tag{2.11}$$

From (2.1) for conic transformation, U is defined by the requirement that

$$\mathsf{U}^{-\mathsf{T}}\Omega_\infty\mathsf{U}^{-1} = \mathsf{I}$$

whence

$$\Omega_\infty = \mathsf{U}^\mathsf{T}\mathsf{U} \tag{2.12}$$

U is in fact the inverse of the internal parameter matrix of the reference camera, as will now be described.

2.3.5 The image of the absolute conic

The absolute conic is imaged in any photograph as a conic dependent only on the internal parameters of the camera. This result demonstrates the similarity invariance of Ω_∞ (its image is independent of the pose of the camera) and can be observed by noting first that for any camera $\mathsf{P} = \mathsf{K}(\mathsf{R}|\ \mathbf{t})$ a point $\mathbf{X} = (\tilde{\mathbf{X}},\ 0)^\mathsf{T}$ on $\boldsymbol{\pi}_\infty$ is imaged as

$$\mathbf{x} = \mathsf{K}(\mathsf{R}|\ \mathbf{t}) \begin{pmatrix} \tilde{\mathbf{X}} \\ 0 \end{pmatrix} = \mathsf{KR}\tilde{\mathbf{X}}$$

Points on the image plane and the plane at infinity are thus related by a plane homography KR. The image of the absolute conic, or IAC, is obtained from transformation of Ω_∞ by (2.1) to give

$$\omega = (\mathsf{KR})^{-\mathsf{T}}\Omega_\infty(\mathsf{KR})^{-1} = \mathsf{K}^{-\mathsf{T}}\mathsf{R}^{-\mathsf{T}}\mathsf{R}^{-1}\mathsf{K}^{-1} = \mathsf{K}^{-\mathsf{T}}\mathsf{K}^{-1}$$

which expands to

$$\omega = \begin{pmatrix} r^2 f^2 & -krf & -rf(rfu_0 - kv_0) \\ -krf & f^2 + k^2 & k(rfu_0 - kv_0) - f^2 v_0 \\ -rf(rfu_0 - kv_0) & k(rfu_0 - kv_0) - f^2 v_0 & r^2 f^4 + f^2 v_0^2 + (rfu_0 - kv_0)^2 \end{pmatrix} \tag{2.13}$$

K can be recovered from ω by Cholesky decomposition[40]. Note that there is an ambiguity in the Cholesky decomposition, since, for any reflection matrix B of the form

$$\mathsf{B} = \begin{pmatrix} \pm 1 & 0 & 0 \\ 0 & \pm 1 & 0 \\ 0 & 0 & \pm 1 \end{pmatrix}, \quad \mathsf{B}^\mathsf{T}\mathsf{B} = \mathsf{I} \text{ and}$$

$\omega = \mathsf{K}^{-\mathsf{T}}\mathsf{K}^{-1} = (\mathsf{B}\mathsf{K}^{-1})^\mathsf{T}\mathsf{B}\mathsf{K}^{-1}$. In practice this sign scaling of any row of K can be chosen to be consistent with the image axes.

The IAC plays an important role in calibration and reconstruction. It encodes the internal parameters of the camera in a form that allows linear constraints to be expressed – a property that will be explored in some depth in this thesis – and does so in a form amenable to geometric manipulation and visualization. Note that is is sometimes convenient to use the dual form of ω: the dual image of the absolute conic or DIAC, ω^{-1}.

The IAC defines the affine transformation $\mathsf{H_a}$ as follows. Given that $\mathsf{P_p} = (\mathsf{I} | \, 0)$, the homography between the plane at infinity and the image plane is the identity. Hence, $\omega = \Omega_\infty$ and, from (2.12)

$$\mathsf{U}^\mathsf{T}\mathsf{U} = \Omega_\infty = \mathsf{K}^{-\mathsf{T}}\mathsf{K}^{-1}$$

and thus the choice of

$$\mathsf{U} = \mathsf{K}^{-1}$$

achieves the objective that $\mathsf{H_a}$ maps the absolute conic to its canonical value. The point transforma-

tion that achieves metric structure from affine is thus

$$\mathsf{H}_\mathsf{a} = \begin{pmatrix} \mathsf{K}^{-1} & \mathbf{0} \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix}$$

2.3.6 Rectified camera matrices and the infinite homography

Projective structure is upgraded to affine by the point transformation H_p. That is, point $\mathbf{X}_\mathsf{a}$ in the affine reconstruction and point $\mathbf{X}_\mathsf{p}$ in the projective reconstruction are related by

$$\mathbf{X}_\mathsf{a} = \mathsf{H}_\mathsf{p}\mathbf{X}_\mathsf{p}$$

Since an image point

$$\mathbf{x} = \mathsf{P}_\mathsf{p}\mathbf{X}_\mathsf{p} = \mathsf{P}_\mathsf{a}\mathbf{X}_\mathsf{a}$$

is the same for both affine and projective camera/reconstruction pairs, the accompanying projection matrices for the affine reconstruction are

$$\mathsf{P}_\mathsf{a} = \mathsf{P}_\mathsf{p}\mathsf{H}_\mathsf{p}^{-1} \quad \text{and} \quad \mathsf{P}'_\mathsf{a} = \mathsf{P}'_\mathsf{p}\mathsf{H}_\mathsf{p}^{-1}$$

Writing $\boldsymbol{\pi}_\infty = (\tilde{\boldsymbol{\pi}}_\infty^\mathsf{T},\ 1)^\mathsf{T}$

$$\mathsf{P}_\mathsf{a} = (\mathsf{I} \mid \mathbf{0}) \begin{pmatrix} \mathsf{I} & \mathbf{0} \\ -\tilde{\boldsymbol{\pi}}_\infty & 1 \end{pmatrix} = \mathsf{P}_\mathsf{p}$$

and

$$\mathsf{P}'_\mathsf{a} = (\mathsf{G} \mid \mathbf{e}') \begin{pmatrix} \mathsf{I} & \mathbf{0} \\ -\tilde{\boldsymbol{\pi}}_\infty & 1 \end{pmatrix} = (\mathsf{G} - \mathbf{e}'\tilde{\boldsymbol{\pi}}_\infty^\mathsf{T} \mid \mathbf{e}') = (\mathsf{H}_\infty \mid \mathbf{e}') \tag{2.14}$$

The left 3×3 sub-matrix of P'_a, H_∞, is known as the infinite homography. H_∞ is the transforma-

tion mapping the projections of ideal points between views: for any ideal 3D point $\mathbf{X} = (\tilde{\mathbf{X}}^\mathsf{T},\ 0)^\mathsf{T}$

$$\mathbf{x} = \mathsf{P_a}\mathbf{X} = \tilde{\mathbf{X}} \ \text{ and } \ \mathbf{x}' = \mathsf{P}'_\mathsf{a}\mathbf{X} = \mathsf{H}_\infty\tilde{\mathbf{X}} = \mathsf{H}_\infty\mathbf{x}$$

The projection matrices for the metric reconstruction follow from the affine as

$$\mathsf{P_m} = \mathsf{P_a}\mathsf{H_a}^{-1} = \mathsf{K}(\mathsf{I}|\ \mathbf{0}) \quad \text{and} \quad \mathsf{P}'_\mathsf{m} = \mathsf{P}'_\mathsf{a}\mathsf{H_a}^{-1} = (\mathsf{H}_\infty\mathsf{K}|\ \mathbf{e}') = \mathsf{K}'(\mathsf{R_c}|\ \mathbf{t_c})$$

with K' the internal parameter matrix of the second camera, $\mathsf{R_c}$ and $\mathbf{t_c}$ the relative rotation and translation of the two cameras and $\mathsf{H}_\infty = \mathsf{K}'\mathsf{R_c}\mathsf{K}^{-1}$.

The general task of uncalibrated metric reconstruction involves finding the eight degrees of freedom of N; three degrees of freedom for $\boldsymbol{\pi}_\infty$ and five for K. Under various conditions on the number and variation in cameras, motion between views and scene geometry, constraints can be obtained on these parameters. The following section describes a number of calibration approaches, ranging from the use of known world geometry to multiple views with arbitrary cameras.

2.4 Calibration and auto-calibration literature

Early camera calibration techniques were designed to identify the calibration parameters to allow measurement of world features or relative three dimensional relationships. The techniques of photogrammetry reflect the immensely detailed calibration necessary to make measurements in such contexts as topographic mapping from aerial photographs [96]. The photogrammetric approach is apparent in early vision calibration methods, where an object with known characteristics is used to infer the internal and external parameters from images. The advent of structure from motion and projective reconstruction without knowledge of camera parameters has made this unnecessary; scene reconstruction is attempted where no calibration objects are present. Auto-calibration, then, is concerned with estimating the camera parameters in order to improve the reconstruction to the level necessary for a task, be it affine or metric. To this end, constraints on the parameters are to be found in the identification of invariant entities, the characteristics of types motion and the relationships between the projections of corresponding image points. A number of approaches are discussed here,

with descriptions divided into four categories:

1. Classical calibration: using the geometry of known objects in the world.

2. Auto-calibration assuming fixed internal parameters: assuming that the same or an identical camera is used for each view of a scene.

3. Auto-calibration with varying internal parameters: relaxing (partially) the assumption of identical cameras.

4. Special or constrained camera motion: solving the special cases where the motion between cameras is known to be of a particular form.

2.4.1 Classical calibration

Classical calibration using known objects originated with photogrammetry, aimed at extremely accurate camera calibration and making use of specially designed calibration devices. The Direct Linear Transformation (DLT) of Abdel-Aziz[1] and the methods of Ganapathy[37] are some of the first attempts to apply the calibration from objects to common types of cameras in a computationally direct manner. The calibration grids of Tsai[107] represent a move towards automated calibration from images. Tsai uses a planar calibration pattern (which has come to be known as the Tsai grid) consisting of a grid of black and white squares. An image of grid points with known world co-ordinates allows calibration from a single view by computing the projection matrix from the images of points and their known world co-ordinates.

The use of cuboid calibration objects exploiting the vanishing points of orthogonal directions appears in the photogrammetry literature of the 1960s, developed by Gracie[41], and introduced to vision by Caprile and Torre[11]. More on this subject appears in chapters 4 and 5, where the method is examined in some detail.

2.4.2 Auto-calibration with constant internal parameters

Auto-calibration was first explored by Faugeras *et al.*[33] using the Kruppa equations. The Kruppa equations embody the constraint that the epipolar planes tangent to the absolute conic project to

corresponding epipolar lines tangent to the IAC in each view. Two independent quadratic constraints on the DIAC can then be written for each view pair, so three views are required to obtain sufficient constraints to solve for the internal parameters. Solution of the resulting six quadratic equations has been approached using continuation[33] and numerical minimization[113], and is in general difficult. However, in the simplified case where the camera has square pixels and a known principal point, a quadratic expression for focal length in terms of the fundamental matrix can be obtained directly from the Kruppa equations[46, 7].

Hartley [44] proposes a method of auto-calibration based on the structure of the projection matrix. The method uses constraints derived from the observation that the QR decomposition of a metric rectified camera must yield an upper triangular K and orthonormal R. An iterative minimization on the parameters of N in (2.9) is performed, with the initialization for the plane at infinity obtained by considering the cheirality of images, the condition that they are the projections of points in front of the camera. Linear programming is used to obtain the estimate from the set of cheiral inequalities.

Heyden and Åström[51] derive a constraint on the DIAC from a formulation that is recognised by Triggs[105] as the projection of the absolute disk quadric:

$$\mathsf{P}_\mathsf{p}\mathsf{Q}_\infty\mathsf{P}_\mathsf{p}^\mathsf{T} = \omega^{-1}$$

Since ω is the same in all views, five non-linear equations on the eight calibration parameters are obtained from each additional camera after the first one.

The *modulus constraint*, developed by Pollefeys *et al.*[81, 80] is an explicit constraint on the plane at infinity. With two identical cameras the infinity homography

$$\mathsf{H}_\infty = \mathsf{K}\mathsf{R}\mathsf{K}^{-1}$$

is conjugate to a rotation and thus has eigenvalues of equal modulus, whence a quartic constraint in the parameters of the plane at infinity is derived. Complete solution is possible from three views, giving the plane at infinity and affine reconstruction, although this requires solving a set of trivari-

ate quartics. A subsequent alternative derivation by Schaffalitzky[89] also classifies the solutions, reducing the number of valid solutions from 64 to 21.

If the infinity homography between two views is known, Luong and Viéville[71] show that the IAC is constrained by the fact that it is the same in both views. Writing

$$\omega_i = \mathsf{H}_\infty^\mathsf{T} \omega_j \mathsf{H}_\infty$$

for the IAC's in views i and j provides four independent linear constraints on the elements of $\omega_i = \omega_j$. This constraint is revisited in chapter 5, where a geometric description of the four constraints is given.

Triggs[106] describes an auto-calibration algorithm from a plane given the homographies between the image of the plane in the first view and its image in all other views. Each homography maps the imaged circular points of the plane to the first view. The imaged circular points lie on the IAC (this will be expanded on in chapter 4), leading to two non-linear constraints on ω from each imaged plane.

2.4.3 Auto-calibration with varying internal parameters

The constant internal parameter limitation of many auto-calibration techniques is impractical for some image sequences. The assumption that a single camera or cameras with identical parameters are used for the entire sequence is often invalid, although some of the parameters may remain constant. It is often reasonable to assume, for example, that a camera has a fixed aspect ratio and zero skew throughout a sequence, but that the focal length and principal point change. This is the approach taken by Heyden and Åström[52], allowing for the varying, unknown focal length and principal point of square pixel cameras. An iterative minimization with the square pixel parametrization of the camera is used to compute metric rectified cameras.

The projection of the absolute disk quadric to the DIAC is used by by Pollefeys *et al.*[79] in an auto-calibration algorithm for square pixel but otherwise varying internal parameters. Assuming initially that the principal point is known, linear constraints on Q_∞ are used to compute an initial estimate from three or more views. A minimization allowing focal length and principal point to

vary follows.

Hartley[43] proposes a method of finding the focal lengths of two cameras for which the other internal parameters are *known.* The method uses a singular value decomposition of the fundamental matrix written in terms of the internal camera parameters.

Nistér[75] modifies the cheirality constraints to compute a reconstruction that is *quasi-affine* with respect to the camera centres rather than the reconstructed points. The quasi-affine reconstruction is important for the subsequent auto-calibration minimization, as a qualitative comparison of methods shows.

2.4.4 Special motion auto-calibration

Certain camera motions, or restrictions on motion, reduce the number of ambiguities in a calibration problem or allow simplified solution. The cases mentioned here include pure translation or rotation of cameras and the planar translation of a camera with rotation about an axis perpendicular to the motion plane.

When a camera is rotating about its optic centre and not translating, all points in a pair of images are related by a homography. Furthermore, this homography is the infinite homography between the cameras. Hartley[45] uses this fact to write constraints on the fixed camera parameters from the invariance of the IAC under transformation by H_∞. Since each view pair provides four independent constraints, a minimum of three allows computation of the internal parameters. De Agapito *et al.*[24] extend this to varying internal parameters by expressing the DIAC in view i in terms of H_∞ and the DIAC in a reference view:

$$\omega_i^{-1} = \mathsf{H}_\infty \omega_0^{-1} \mathsf{H}_\infty^\mathsf{T} \tag{2.15}$$

Iterative minimization is then used to compute the cameras in all views. A further extension[23] returns to the IAC form of (2.15) and obtains linear constraints on the IACs from partial knowledge of internal parameters. Zero skew induces a constraint from the zero entry in ω, and similar constraints follow from a known aspect ratio or principal point. This approach has been generalized to include translational camera motions by Hartley *et al.*[50]. Cheirality constraints are used to provide an

initial estimate of the infinite homography, and constraints from the transfer of the IAC provide a measure of error for subsequent minimization.

For a pair of cameras with a pure translation between them, the infinite homography is the identity matrix. Accordingly, Moons *et al.*[72] show that for a sequence in which the first motion is a pure translation, affine structure can be found directly. Armstrong *et al.*[2] use the requirement of pure translation followed by rotation to obtain metric structure. Pajdla and Hlaváč[76] uses three translations of a camera to obtain metric structure.

Auto-calibration schemes exist for the case of planar motion and rotation orthogonal to the translation plane (typically in the context of robot navigation). Under such motions, a fixed line (the vanishing line of the translation plane) with fixed imaged circular points and a line of fixed points (the image of the screw axis) exist for view pairs. Since these fixed entities are images of ideal entities, their image co-ordinates constrain the plane at infinity. Armstrong *et al.*[3] use three views to compute the fixed entities and hence the plane at infinity and affine structure. Beardsley and Zisserman [5] demonstrate affine reconstruction with a *stereo rig* undergoing planar motion. In this case the mapping between the two view projective reconstructions before and after the motion induce constraints on the fixed entities. A general stereo calibration scheme using the fixed entities can be found in Zisserman *et al.*[116] and Devernay and Faugeras[28].

2.4.5 Where this thesis fits in

The methods presented in subsequent chapters combine elements of calibration objects, knowledge of some internal parameters and auto-calibration constraints.

Much of the material turns familiar world objects into calibration objects. In chapter 3 it will be seen how scene plane geometry can be exploited to achieve metric reconstruction of the plane, as well as provide constraints on the internal parameters of the camera. Chapter 4 extends the notion of the vanishing points of orthogonal directions providing constraints on the camera, combining it with the square pixel camera model. Constraints are obtained from images of structured environments, such as buildings, where the right angles and parallel lines that dominate human constructions provide calibration information.

In two views (chapter 5) the presence of matched vanishing points defines the plane at infinity. Auto-calibration methods of computing the internal parameters given the infinite homography can thus be used in addition to the single view constraints imposed by the vanishing points. Specifically, the infinite homography will be shown to transfer the constraints obtained from knowing that the cameras have square pixels between views.

Even the human body will be used to calibrate affine cameras – the fixed lengths of parts of the body as it moves provide the constraints in chapter 6.

Chapter 3

Perspective Correction of Images of Planes

3.1 Introduction: From perspective distortion to fronto-parallel images of planes

The image of a world plane captured by a camera experiences perspective distortion. Figure 3.1 shows such an image of a plane, a photograph of an external wall of a building. Observe that rectangular features, such as the windows, do not appear as rectangles in the image: the plane geometry has undergone a transformation that distorts Euclidean properties such as length, angle and parallelism. Figure 3.2 shows the perspective corrected wall, where the plane geometry is that which would be seen had the original photograph been taken with the camera fronto-parallel to the wall. Such perspective correction or *rectification* is the subject of this chapter.

Perspective correction has a number of applications, including image mosaicing [100] and planar object recognition [87]. More topically, rectification is essential for 3D measurement and reconstruction of objects including planes, a subject which will be dealt with in the following chapters.

The nature of plane perspective distortion is well understood – the map between the world plane and the perspective image is a plane projective transformation, a 2D homography[36, 91, 101]. Correcting perspective distortion is a matter of determining the eight degrees of freedom of the

Figure 3.1: A photograph of an external wall of Keble College, Oxford. The perspective distortion of the wall plane is clear from the shape of rectangular structures, such as the windows.

Figure 3.2: The perspective distortion *corrected* in the view of the wall of Keble College. Such a view would be obtained from a fronto-parallel photograph.

homography. The parameters of the homography depend on the relative position and orientation of the world plane with respect to the camera, and the camera internal parameters. Often, however, very little or none of this information is available.

Collins and Beveridge [15] developed a technique of partial plane rectification. They showed that once the vanishing line of a plane is identified, the transformation from world to image plane can be reduced to an affine transformation. This result was used in registering satellite images, reducing the dimension of the search space from eight, for a full projective homography, to six, the degrees of freedom of an affine transformation. They showed further that the orientation of the plane relative to the camera can be computed *provided the internal calibration is known.*

The work presented in this chapter goes further, reducing the world to image plane homography to a similarity transformation. Metric measurements may then be made on the plane. The *rectifying homography* for the plane is computed from scene geometric information, specifically parallelism of lines, the angles between lines and ratios of lengths along lines in different directions. Later it will be shown that a rectified plane provides constraints on unknown internal camera parameters. Conversely, a known internal calibration provides constraints on the plane rectification homography.

The chapter proceeds by examining the rectification homography more closely and decomposing it into metric and non-metric parts. The metric part is then further decomposed into two parts associated with geometric features, the vanishing line of the plane and the imaged circular points. Various methods of computing the rectification are described.

The intersection of parallel world lines in a vanishing point is a common computation in these techniques, as well as in the work of later chapters. Consequently, a method is presented to compute a Maximum Likelihood Estimate (MLE) of the intersection of a number of (noisy) measured lines. Finally, a first order error analysis of line intersection is performed and tested.

3.2 Rectification geometry

3.2.1 The rectification homography

The rectification homography that maps a point on the image plane, $\mathbf{x}$, to a point on the world plane, $\mathbf{x}'$, is represented by a 3×3 homogeneous matrix H such that

$$\mathbf{x}' = \mathsf{H}\mathbf{x} \tag{3.1}$$

H has eight degrees of freedom.

Just as for the 3D homography rectifying a projective reconstruction, H may be decomposed as [37]

$$\mathsf{H} = \mathsf{M}\ \mathsf{N} \tag{3.2}$$

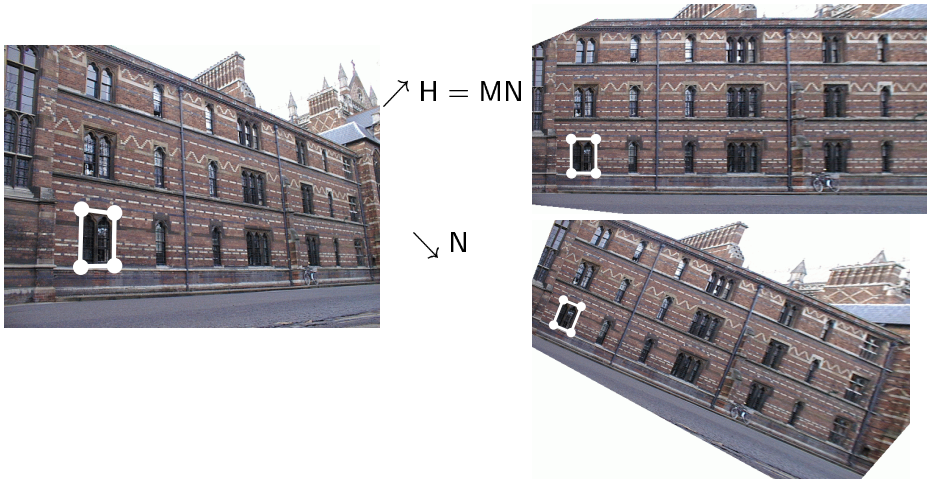
The transformation M is the *metric* part of the homography and is a similarity transformation, which can be a reflection or of the form

$$\mathsf{M} = \begin{pmatrix} s\mathsf{R} & \mathbf{t} \\ \mathbf{0}^{\mathsf{T}} & 1 \end{pmatrix} \tag{3.3}$$

where R is a rotation matrix, $\mathbf{t}$ a translation vector, and s an isotropic scaling. There are four degrees of freedom in M.

The remaining four degrees of freedom in the rectifying homography are encoded by N, the non-metric component of H. N has affine and projective components which will be described in the next section.

The established method of computing H is from the correspondence of four (or more) points *with known position*[20]. Given four points (3.1) may be written out explicitly for known $\mathbf{x}$ and $\mathbf{x}'$. The four points give eight equations in the eight degrees of freedom of H and the elements of H may then be computed. Once H is determined the image can be warped onto the world plane and in this way a rectified image is obtained.



(a) (b)

Figure 3.3: Two metric rectifications of Keble wall plane in (a) appear in (b). The metric rectifications differ only by a similarity transformation, a rotation, translation and scaling of the plane co-ordinate system. Metric properties on the plane are the same.

However, it is not necessary to determine the entire homography in order to obtain a *metric* rectification. The metric properties of the plane, such as angle and relative length, are invariant to M since it is a similarity transformation. The complete metric rectification is thus known when N is computed.

Consider, for example, figure 3.3(a). The dimensions of the rectangle defined by the four corners of the window shown are known. The four corners of the rectangle thus have known world co-ordinates as soon as a co-ordinate frame on the world plane, that is, the building wall, is chosen. The two images in figure 3.3(b) show rectifications for two different choices of world co-ordinate frame. The two rectifications differ only by a similarity transformation.

It will be shown in the following sections that the metric rectification can be computed without the known co-ordinates of four points. This is done by parametrizing N by plane geometric invariants that can be computed from geometric features such as parallelism and angle.

3.2.2 The non-metric part of H

The non-metric part of H can be decomposed into two matrices:

$$\mathsf{N} = \mathsf{H}_\mathsf{a}\mathsf{H}_\mathsf{p} \tag{3.4}$$

Four of the eight degrees of freedom of H have been discarded with M, and it will now be shown how the remaining four degrees of freedom of N are parametrized by H_a and H_p and associated with plane geometry. This in turn leads to methods for computing N from scene geometry.

Specializations of the image plane

The world plane is a Euclidean plane – all the Euclidean properties such as length and angle are defined on this plane. The line at infinity has the canonical co-ordinates $(0,\ 0,\ 1)^\mathsf{T}$ and the circular points are $(1,\ \pm i,\ 0)^\mathsf{T}$.

In the absence of further knowledge, the image plane is a projective plane. Nothing may be said about Euclidean or affine properties of the image plane geometry. However, as soon as a line at infinity $\mathbf{l}_\infty$ on the image place is chosen affine properties are defined. The choice of a pair of circular points $\mathbf{I}$ and $\mathbf{J}$ for the image plane in turn defines metric properties. Furthermore, it is possible to measure both affine and metric properties. For example, the affine property of parallelism is defined since parallel lines intersect on $\mathbf{l}_\infty$. The metric property of angle between lines can be computed by Laguerre's formula from the cross ratio of the circular points and the intersections of the lines with $\mathbf{l}_\infty$.

Now, from (3.1) the world and image planes are projectively equivalent. The key idea here is that if the particular line at infinity and circular points chosen on the image plane are images of the corresponding canonical elements on the world plane, the affine and metric properties of the world plane can be measured on the image plane.

So, assume it is possible to identify the projections of the canonical world plane line at infinity and circular points. Selecting these as $\mathbf{l}_\infty$, $\mathbf{I}$ and $\mathbf{J}$, the affine and metric properties of the image plane are identical to those of the world plane. The metric rectification homography N can now be

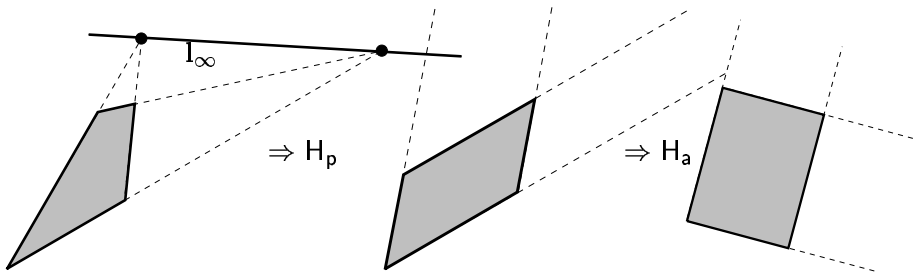


Figure 3.4: A projectively distorted rectangle on the left is mapped to a parallelogram by H_p, the transformation that maps $\mathbf{l}_\infty$ 'to infinity', to the Euclidean line at infinity $(0,\ 0,\ 1)^\mathsf{T}$). The parallelogram is in turn mapped to a rectangle by H_a, the transformation that maps the circular points to their canonical values.

seen to be a homography that maps these choices of $\mathbf{l}_\infty$, $\mathbf{I}$ and $\mathbf{J}$ to their canonical values *so that measurements on image plane co-ordinates can be transformed to agree with familiar Euclidean algebra*. That is, N maps $\mathbf{I}$ and $\mathbf{J}$ to $(1,\ \pm i,\ 0)^\mathsf{T}$. The decomposition of N in (3.4) is a stratification of the transformation such that H_p maps $\mathbf{l}_\infty$ to $(0,\ 0,\ 1)^\mathsf{T}$ and H_a maps $\mathsf{H}_\mathsf{p}\mathbf{I}$ and $\mathsf{H}_\mathsf{p}\mathbf{J}$ to $(1,\ \pm i,\ 0)^\mathsf{T}$.

Figure 3.4 shows the effect of transformations H_p and H_a on a projectively distorted rectangle.

Parametrization of H_a and H_p

H_p is required to map $\mathbf{l}_\infty$ to the line $(0,\ 0,\ 1)^\mathsf{T}$. Writing $\mathsf{H}_\mathsf{p} = (\mathbf{p}_1\ \mathbf{p}_2\ \mathbf{p}_3)^\mathsf{T}$ and transforming the line by

$$\mathbf{l}_\infty = \mathsf{H}_\mathsf{p}^\mathsf{T}(0,\ 0,\ 1)^\mathsf{T} = \begin{pmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \mathbf{p}_3$$

So H_p acts as required if $\mathbf{p}_3 = \mathbf{l}_\infty = (l_{\infty 1},\ l_{\infty 2},\ l_{\infty 3})^\mathsf{T}$. The rows $\mathbf{p}_1$ and $\mathbf{p}_2$ can be chosen so that H_p has the form

$$\mathsf{H}_\mathsf{p} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ l_{\infty 1} & l_{\infty 2} & l_{\infty 3} \end{pmatrix} \tag{3.5}$$

H_p has two degrees of freedom since $\mathbf{l}_\infty$ is a homogeneous vector.

The circular points lie on the line at infinity. Following transformation by H_p, the circular points of the affine image plane $\mathsf{H}_\mathsf{p}\mathbf{I}$ and $\mathsf{H}_\mathsf{p}\mathbf{J}$ lie on the canonical line at infinity and thus have third co-ordinate zero. The circular points can then be written as

$$\begin{aligned} \mathbf{I}_\mathsf{a} &= \mathsf{H}_\mathsf{p}\mathbf{I} = (\alpha - i\beta,\ 1,\ 0)^\mathsf{T} \text{and} \\ \mathbf{J}_\mathsf{a} &= \mathsf{H}_\mathsf{p}\mathbf{J} = (\alpha + i\beta,\ 1,\ 0)^\mathsf{T} = \overline{\mathbf{I}}_\mathsf{a} \end{aligned} \tag{3.6}$$

$\mathbf{I}_\mathsf{a}$ and $\mathbf{J}_\mathsf{a}$ may also be treated as homogeneous points on the projective line by simply neglecting the third co-ordinate: $\mathbf{p_I} = (\alpha - i\beta,\ 1)^\mathsf{T}$. The circular points may then also be be normalised and written as a scalar pair

$$I = \alpha - i\beta \text{ and } J = \alpha + i\beta$$

The parameters α and β are simply the real and imaginary parts of I and J.

H_a is required to map $\mathbf{I}_\mathsf{a}$ to $(1,\ i,\ 0)^\mathsf{T}$. Sparing the reader the details, by writing out $\mathsf{H}_\mathsf{a}\mathbf{I}_\mathsf{a} = (1,\ i,\ 0)^\mathsf{T}$ and separating real and imaginary parts H_a can be shown to have the form

$$\mathsf{H}_\mathsf{a} = \begin{pmatrix} \frac{1}{\beta} & -\frac{\alpha}{\beta} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{3.7}$$

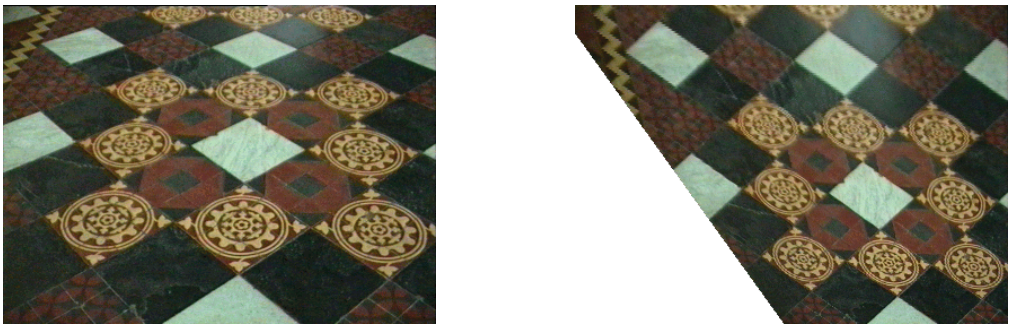
The parameters α and β of H_a account for the final two degrees of freedom of N. Note that there remains some ambiguity on the metric image plane since the circular points cannot be distinguished. Swapping $\mathbf{I}_\mathsf{a}$ and $\mathbf{J}_\mathsf{a}$ or equivalently interchanging the first two rows of H_a results in a *reflection* of the metric image plane.

The sequential application of H_p and H_a to the image plane is a *stratified rectification* of the plane. It is analogous to the stratified rectification of a 3D projective reconstruction described in chapter 1. An unstratified approach to plane rectification, in which the circular points are identified without explicitly computing $\mathbf{l}_\infty$ will appear in section 3.5

3.3 Computing stratified metric rectifications

3.3.1 From projective to affine

The first stage is to determine H_p, which requires identifying the vanishing line $\mathbf{l}_\infty$. Once H_p is computed the image geometry can be transformed to affine rectify the image. Image texture may also be warped to produce a view of the plane, as in figure 3.5.



(a) (b)

Figure 3.5: The floor of the chapel of Merton College, Oxford. (a) A photograph of the floor. There are a number of square tiles and circle patterns. (b) The affine rectified floor plane after transformation by H_p. Note that the parallelism of the tile boundary lines is restored, but the angles between lines and relative lengths do not appear correct. Note that for aesthetic reasons a rotation has been applied to the initial image before transformation. This has no effect on the affine properties.

The property that parallel lines intersect on $\mathbf{l}_\infty$ provides a convenient method of computing H_p. A set of parallel lines intersect in a vanishing point in the image, with $\mathbf{l}_\infty$ then the line of vanishing points. Given two vanishing points $\mathbf{v}_1$ and $\mathbf{v}_2$

$$\mathbf{l}_\infty = \mathbf{v}_1 \times \mathbf{v}_2$$

If more than two parallel lines in a particular direction are present, the vanishing point is over-constrained. A Maximum Likelihood Estimate (MLE) of such a vanishing point is described in section 3.6. An example of sets of parallel lines in two directions, showing the vanishing points and vanishing line for the example of figure 3.5 above, appears in figure 3.6.

A second method of determining a vanishing point can be found from a known ratio of lengths from a reference in one direction [44]. Consider the configuration shown in figure 3.7, where a line

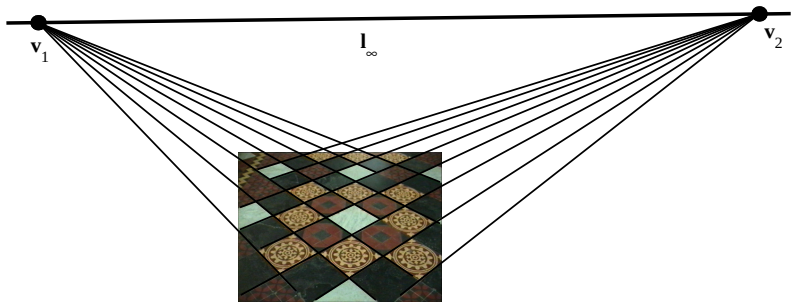


Figure 3.6: Two sets of parallel lines intersect in vanishing points $\mathbf{v}_1$ and $\mathbf{v}_2$. The vanishing points define $\mathbf{l}_\infty = (0.0121,\ 0.9999,\ 500.4788)^\mathsf{T}$.

$\mathbf{l}$ undergoes a projective transformation to $\mathbf{l}'$. The point from which distances are measured is taken as the origin and, with the two known distances and the point at infinity, these define a cross ratio of four points, which is a projective invariant. As such, it has the same value after transformation and the vanishing point must be in the same relationship after the projectivity as before. It will be shown here that the vanishing point can be calculated from the known ratio of lengths along the line by considering the homography of the line directly as well as by writing out the cross ratio.

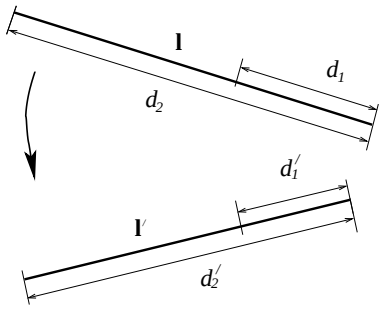


Figure 3.7: Notation for computing the vanishing point from a known ratio of lengths long a line.

Treating the projective transformation of points on the line as a homography on $\mathbb{P}^1$ taking point

$\mathbf{p}$ to $\mathbf{p}'$:

$$\mathbf{p}' = \begin{pmatrix} h_1 & h_3 \\ h_2 & 1 \end{pmatrix} \mathbf{p}$$

Taking points $(0,\ 1)^\mathsf{T}$ at the origin, $\mathbf{p}_1 = (d_1,\ 1)^\mathsf{T}$ and $\mathbf{p}_2 = (d_2,\ 1)^\mathsf{T}$, and regarding the origin as fixed implies that $h_3 = 0$.

The point at infinity, $\mathbf{p}_\infty = (1,\ 0)^\mathsf{T}$ maps to $\mathbf{p}'_\infty = (h_1,\ h_2)^\mathsf{T}$, so the distance of the vanishing point from the origin is $v = h_1/h_2$. It is necessary, then, to determine the ratio h_1/h_2 to find the vanishing point.

Points $\mathbf{p}'_1$ and $\mathbf{p}'_2$ in inhomogeneous form are given by

$$\begin{aligned} d'_1 &= \frac{h_1 d_1}{h_2 d_1 + 1} \\ d'_2 &= \frac{h_1 d_2}{h_2 d_2 + 1} \end{aligned}$$

being two equations in h_1 and h_2. Solving simultaneously gives

$$v = \frac{h_1}{h_2} = \frac{d'_1 d'_2 (\frac{d_1}{d_2} - 1)}{\frac{d_1}{d_2} d'_2 - d'_1} \tag{3.8}$$

Only the ratio of original lengths d_1/d_2 is required. This result can also be obtained directly from the cross ratio by considering

$$\mathfrak{C}\{d_1, d_2; 0, \infty\} = \mathfrak{C}\{d'_1, d'_2; 0, v\}$$

Writing out the cross ratios gives

$$\frac{d_1}{d_2} = \frac{d'_1}{d'_2} \cdot \frac{d'_2 - v}{d'_1 - v}$$

and

$$v = \frac{d'_1 d'_2 (\frac{d_1}{d_2} - 1)}{\frac{d_1}{d_2} d'_2 - d'_1}$$

An example of computing vanishing points from a length ratio appears in figure 3.8. The equal length ratios provided by the sides of the square tiles provide the two vanishing points. The results are much the same as figure 3.6.

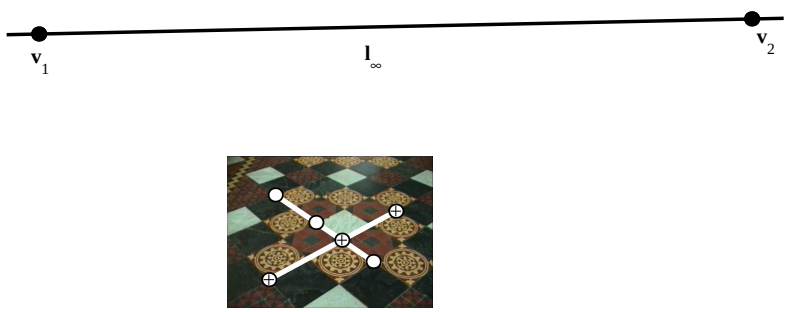


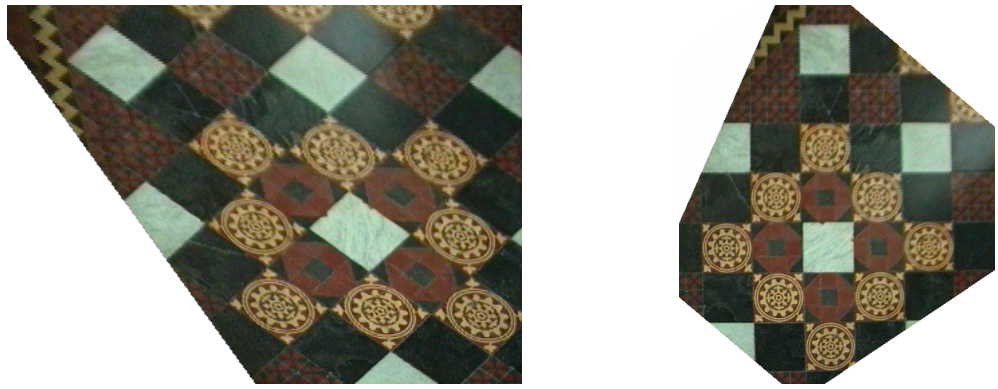
Figure 3.8: Computing the vanishing line from length ratios. The square tiles provide relative length information: vanishing point $\mathbf{v}_1$ is computed from the ratio of lengths of the points shown as circles, and $\mathbf{v}_2$ computed from the ratio of lengths of the points depicted by crosses. The vanishing points define $\mathbf{l}_\infty = (0.0212,\ 0.9998,\ 473.1308)^\mathsf{T}$. Compare this to figure 3.6.

Other constraints may be used to determine the vanishing line. For example, a single set of equally spaced parallel lines on the plane is sufficient to determine $\mathbf{l}_\infty$[90].

3.3.2 From affine to metric

Having recovered the plane geometry up to an affine transformation by applying the matrix H_p, the final stage is the recovery of metric geometry. This requires an affine transformation of the plane, H_a, that will restore angles and length ratios for non-parallel lines. The goal is thus to determine α and β, the real and imaginary parts of $\mathbf{I}$ and $\mathbf{J}$. This fully defines H_a. Once H_a is known, the affine rectified plane can be metric rectified, as in figure 3.9.

Three constraints on the circular points of the affine rectified plane are derived below. The



(a) (b)

Figure 3.9: Metric rectification from affine of the floor of the chapel of Merton College. (a) Affine rectification of the plane. (b) Metric rectified plane. Observe the square floor tiles and circle patterns.

constraints are quadratic in α and β and may be represented as circles in the complex $(\alpha,\ \beta)$ plane, allowing solution by simple intersection of circles. The constraints are found from

1. known angles between lines,
2. equality of unknown angles, and
3. known ratios of lengths.

Known angles

Consider two lines $\mathbf{l}_a$ and $\mathbf{l}_b$ transformed to lines $\mathbf{l}'_a$ and $\mathbf{l}'_b$ by some affine transformation $\tilde{\mathsf{H}}_{\mathsf{a}}$. Angle θ between $\mathbf{l}_a$ and $\mathbf{l}_b$ is known, but the directions of these two lines are not. The intersections of a line $\mathbf{l} = (l_1,\ l_2,\ l_3)^\mathsf{T}$ with $\mathbf{l}_\infty$ is the point $(-l_2,\ l_1,\ 0)^\mathsf{T}$, which can be considered as a point along the one dimensional line, represented non-homogeneously as $-l_2/l_1$. The intersections of $\mathbf{l}'_a$ and $\mathbf{l}'_b$ with $\mathbf{l}_\infty$ are points denoted a and b. Taking the cross ratio of a and b with I and J, Laguerre's formula gives

$$\mathfrak{C}\{a, b; I, J\} = \cos 2\theta + i \sin 2\theta$$

Writing out the cross ratio and substituting $\alpha - i\beta$ and $\alpha + i\beta$ for I and J,

$$\frac{a-(\alpha-i\beta)}{b-(\alpha-i\beta)} \cdot \frac{b-(\alpha+i\beta)}{a-(\alpha+i\beta)} = \cos 2\theta + i \sin 2\theta$$

Multiplying through and equating real parts gives

$$\alpha^2 - (a+b)\alpha + \beta^2 - \cot\theta(a-b)\beta + ab = 0 \tag{3.9}$$

which, by completing the square terms in α and β, represents a circle in the complex plane with centre

$$(c_\alpha, c_\beta) = (\frac{(a+b)}{2}, \frac{\cot\theta(a-b)}{2})$$

and radius

$$r = \|\frac{(a-b)}{2\sin\theta}\|$$

Equating imaginary parts yields the identical relationship.

The complex plane circle represents a family of solutions for the circular points. In the case of a right angle, the circle is centred on the real axis and passes through both complex conjugate points I and J. Circles obtained from other angles move off the real axis, passing through one of the circular points, in which case the complementary angle between the same pair of lines results in a circle through the complex conjugate point. The right angle case is by far the most important since orthogonality is ubiquitous in human constructions. The presence of orthogonality alone also simplifies the parametrization of $\mathsf{H_a}$ – an algorithm for the case of buildings is presented in the following section.

Equal unknown angles

An additional constraint can be found from the knowledge of two angles of the same magnitude. Taking (3.9) for *two pairs* of lines with points on $\mathbf{l}_\infty$ at a_1, b_1, a_2 and b_2, and solving yields

$$\beta \cdot \cot\theta = \frac{-a_1 - b_1 + a_2 + b_2}{a_1 - b_1 - a_2 + b_2}\alpha + \frac{-a_1 - b_1 + a_2 + b_2}{a_1 b_1 - a_2 b_2}$$

Substitution into one of the initial equations defines a circle centred on the real axis at

$$(c_\alpha, c_\beta) = (\frac{a_1 b_2 - b_1 a_2}{a_1 - b_1 - a_2 + b_2}, 0)$$

with squared radius

$$r^2 = (\frac{a_1 b_2 - b_1 a_2}{a_1 - b_1 - a_2 + b_2})^2 + \frac{(a_1 - b_1)(a_1 b_1 - a_2 b_2)}{a_1 - b_1 - a_2 + b_2} - a_1 b_1$$

Known length ratio

Consider two lines with lengths $\mathbf{l}_1$ and $\mathbf{l}_2$ which have endpoints as shown in figure 3.10, and lengths d_1 and d_2 which have known ratio $s = d_1/d_2$. The metric lengths are related by

$$(x_{11} - x_{12})^2 + (y_{11} - y_{12})^2 = s^2((x_{21} - x_{22})^2 + (y_{21} - y_{22})^2)$$

Transforming the endpoints by $\mathsf{H_a}$, this can be expressed as

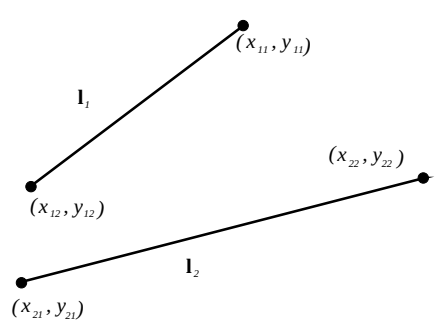


Figure 3.10: Affine constraint from known length ratio.

$$(\frac{-x'_{11}}{\beta} + \frac{\alpha y'_{11}}{\beta} + \frac{x'_{12}}{\beta} - \frac{\alpha y'_{12}}{\beta})^2 + (y'_{11} - y'_{12})^2$$
$$= s^2((\frac{-x'_{21}}{\beta} + \frac{\alpha y'_{21}}{\beta} + \frac{x'_{22}}{\beta} - \frac{\alpha y'_{22}}{\beta})^2 + (y'_{21} - y'_{22})^2)$$

which, writing $\Delta x'_n$ for $x'_{n1} - x'_{n2}$ and similarly for y', becomes

$$(-\frac{1}{\beta}\Delta x'_1 + \frac{\alpha}{\beta}\Delta y'_1)^2 + \Delta y'^2_1 = s^2((-\frac{1}{\beta}\Delta x'_2 + \frac{\alpha}{\beta}\Delta y'_2)^2 + \Delta y'^2_2)$$

Squaring, rearranging and multiplying through by β^2 gives

$$(\Delta y'^2_1 - s^2\Delta y'^2_2)\alpha^2 - 2(\Delta x'_1\Delta y'_1 - s^2\Delta x'_2\Delta y'_2)\alpha +$$
$$(\Delta y'^2_1 - s^2\Delta y'^2_2)\beta^2 + \Delta x'^2_1 - s^2\Delta x'^2_2 = 0$$

which is a circle with centre

$$(c_\alpha, c_\beta) = (\frac{(\Delta x'_1\Delta y'_1 - s^2\Delta x'_2\Delta y'_2}{\Delta y'^2_1 - s^2\Delta y'^2_2}, 0)$$

and radius

$$r = \|\frac{s(\Delta x'_2\Delta y'_1 - \Delta x'_1\Delta y'_2)}{\Delta y'^2_1 - s^2\Delta y'^2_2}\|$$

Note also that this constraint can be linearized by over parametrizing the affine rectification.[1] The squared metric length of a line segment $\Delta\mathbf{x}' = \mathsf{H_a}\Delta\mathbf{x}$ is its inner product

$$d^2 = (\mathsf{H_a}\Delta\mathbf{x})^\mathsf{T}\mathsf{H_a}\Delta\mathbf{x} = \Delta\mathbf{x}^\mathsf{T}(\mathsf{H_a^T}\mathsf{H_a})\Delta\mathbf{x}$$

Given two line segments $\Delta\mathbf{x}_1$ and $\Delta\mathbf{x}_2$, with a metric length ratio s, the constraint

$$\Delta\mathbf{x}_1^\mathsf{T}(\mathsf{H_a^T}\mathsf{H_a})\Delta\mathbf{x}_1 = s^2\Delta\mathbf{x}_2^\mathsf{T}(\mathsf{H_a^T}\mathsf{H_a})\Delta\mathbf{x}_2$$

[1] This was pointed out by Dr Stefan Carlsson in the 3D affine reconstruction context, and forms the basis of the 3D affine rectification constraint in chapter 6.

provides a single linear constraint on the elements of

$$\mathsf{H}_{\mathsf{a}}^{\mathsf{T}}\mathsf{H}_{\mathsf{a}} = \begin{pmatrix} 1 & -\alpha & 0 \\ -\alpha & \alpha^2+\beta^2 & 0 \\ 0 & 0 & \beta^2 \end{pmatrix}$$

In the examples that follow, the quadratic form is used, since this allows direct combination of the different types of constraints on α and β.

Example constraints

The three types of constraints described above are demonstrated in figure 3.11. Two pairs of orthogonal lines provide two known angle constraints, the equality of the angles between the lines provides a third constraint and a fourth is computed from the ratio of length of the sides of a square tile. The metric rectified image plane shown in figure 3.9 is obtained from these constraints. The inevitable presence of image noise means that the circles do not all intersect in a single point, so the affine parameters are estimated by computing the point on the upper half of the $(\alpha,\ \beta)$ plane closest to all the circles. A few comments are in order at this stage:

1. All three constraints are functions of line *direction* rather than line position. Parallel lines thus provide identical constraints in the absence of noise.

2. Consistent constraints can be obtained from different methods. For example, the constraint that the sides of a square have the same length and the constraint that the diagonals of the square are orthogonal are identical.

3. As a fundamental property, the circular points can also be computed directly from the intersection of the line at infinity and a circle on the plane.

3.4 The case of buildings

Rectification from scene geometry is particularly useful in the reconstruction of buildings. Such reconstructions commonly require rectification of building facades where rectangular structures exist,

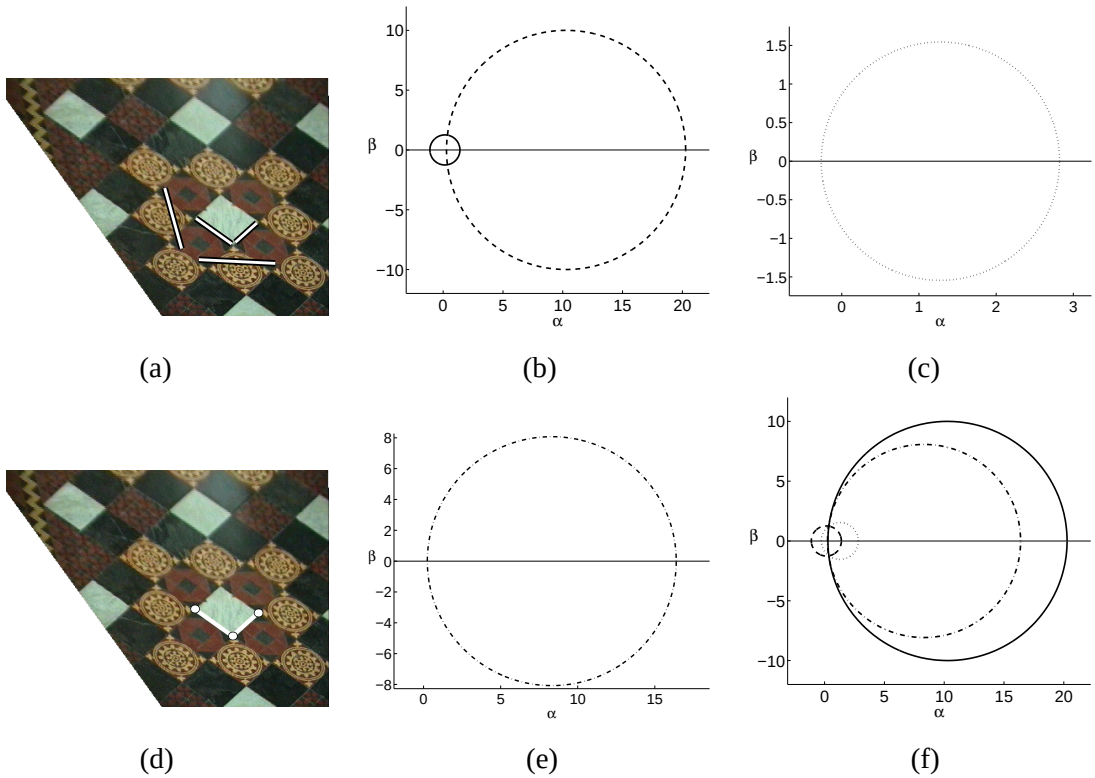


Figure 3.11: Constraints on the affine rectification parameters. (a) Two pairs of line segments from which constraints are obtained in two ways: (b) two constraints defined by the right angles between the pairs and (c) a single constraint from the *equality* of the angles between the two line pairs. (d) The length ratio of the sides of the square provides a single constraint, shown in (e). All the constraints are combined in (f). The computed affine parameters are $\alpha = 0.3558$ and $\beta = 1.2388$. These parameters are responsible for the metric rectification of figure 3.9.

such as the facade outline or windows (as in figure 3.1). The typical structure presents orthogonal sets of parallel lines in the vertical and horizontal directions. The vertical and horizontal parallel lines define $\mathbf{l}_\infty$ and the affine rectification of the plane (see figure 3.12).

The orthogonality of the vertical and horizontal directions provides a single constraint on the affine circular points and thus on the affine rectification. A one parameter family of affine transformations remains – the ambiguity corresponding to the relative scale of vertical and horizontal directions. Figure 3.13 shows two of the possible rectifications of a building facade. The first is computed with an arbitrary selection of relative scale of a window on the facade. The second shows the rectification from the beginning of the chapter, where the aspect ratio of a window is known from measurement and the ambiguity is thus removed. A parametrization of the rectification homography

that makes this ambiguity explicit will now be described.

(a) (b)

Figure 3.12: A typical building facades has two pairs of parallel lines in orthogonal directions. The parallel lines determine $\mathbf{l}_\infty$ and thus the affine rectification of the plane.

(a) (b)

Figure 3.13: The aspect ratio ambiguity in relative scale of vertical and horizontal directions. (a) An incorrectly scaled image. (b) The correctly scaled image, from the known length ratio of the sides of a window.

The previous sections have described the stratified rectification of a plane in two stages; first a projective and then an affine transformation. In the case of a pair of orthogonal directions and the aspect ratio in these directions it is possible to further stratify the rectification by decomposing H_a as

$$\mathsf{H}_\mathsf{a} = \mathsf{A}_2\mathsf{A}_1\mathsf{R}_\mathsf{a}$$

The three stages of the decomposition have the purposes of

1. R_a: Rotate a direction to the horizontal axis.

2. A_1: Transform a second direction to the vertical axis without changing the orientation of the horizontal axis.

3. A_2: Set the aspect ratio of vertical and horizontal directions without changing the direction of either.

To parametrize A_2, A_1 and R_a, consider two orthogonal vanishing points v_1 and v_2. H_p is computed according to (3.5). The effect of H_p on the vanishing points is to transform them to the form $\mathbf{v}_{1\mathsf{A}} = (\times,\ \times,\ 0)^\mathsf{T}$, where they define *directions.* $\mathbf{v}_{1\mathsf{A}}$ can can be written as a unit norm direction vector $\mathbf{v}_{1\mathsf{A}} = (\cos(\phi),\ \sin(\phi),\ 0)^\mathsf{T}$, where ϕ is the angle $\mathbf{v}_{1\mathsf{A}}$ makes with the horizontal axis. Now R_a is a rotation matrix that rotates $\mathbf{v}_{1\mathsf{A}}$ to the horizontal axis: $\mathsf{R}\mathbf{v}_{1\mathsf{A}} = (1,\ 0,\ 0)^\mathsf{T}$:

$$\mathsf{R}_\mathsf{a} = \begin{pmatrix} \cos(\phi) & \sin(\phi) & 0 \\ -\sin(\phi) & \cos(\phi) & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

If the angle between $\mathbf{v}_{1\mathsf{A}}$ and $\mathbf{v}_{2\mathsf{A}}$ is θ, $\mathbf{v}_{2\mathsf{A}}$ now makes an angle of $\pi - \theta$ with the vertical axis and the transformation

$$\mathsf{A}_1 = \begin{pmatrix} 1 & -\cot(\theta) & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

meets criterion 2 above.

The final transformation A_2 is simply

$$\mathsf{A}_2 = \begin{pmatrix} \mu & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

where μ corrects the relative scale in horizontal and vertical directions.

The relative scaling μ is the unknown parameter defining the ambiguity above. It may be determined from measurement, as in the example, but it is also possible to compute it using the internal parameters of the camera. In chapter 4 it will be shown how the relationship between plane rectification and camera internal parameters can be used to resolve the ambiguity *without* measurement.

3.5 Unstratified rectification

The stratified approach to plane rectification considered thus far in this chapter separates the projective and affine components of the rectification homography. This is a useful feature – it is possible to combine constraints from a range of geometric characteristics to compute the projective and affine parameters. Of the four rectification parameters, two are associated with l_∞ and two with $\mathbf{I}_\mathsf{a}$ and $\mathbf{J}_\mathsf{a}$, the circular points on the canonical line at infinity. However, it is also possible to consider the circular points on the projective image plane, $\mathbf{I}$ and $\mathbf{J}$. Together, $\mathbf{I}$ and $\mathbf{J}$ encapsulate all four degrees of freedom of the rectification homography. The need to combine them in order to perform useful calculations leads to a *dual conic* representation of the point pair. This conic is a 2D analogue of the absolute disk quadric, encoding both projective and affine specializations of the plane in a single entity, just as the absolute disk quadric does in 3D. The conic and its parametrization will be described first, followed by some properties and finally a method to compute the circular points, and thus the rectification homography, from orthogonal lines on the plane.

3.5.1 The conic dual to the circular points

A conic D_∞ dual to the circular points can be defined as

$$\mathsf{D}_\infty = \mathbf{I}\mathbf{J}^\mathsf{T} + \mathbf{J}\mathbf{I}^\mathsf{T} \tag{3.10}$$

On the metric plane (denoting metric entities with primes) $\mathbf{I}' = (1,\ i,\ 0)^\mathsf{T}$ and $\mathbf{J}' = (1,\ -i,\ 0)^\mathsf{T}$, and

$$\mathsf{D}'_\infty = \mathbf{I}'\mathbf{J}'^\mathsf{T} + \mathbf{J}'\mathbf{I}'^\mathsf{T} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{3.11}$$

On the image plane the circular points and D_∞ can be parametrized by transformation of $\mathbf{I}'$ and $\mathbf{J}'$

from the metric plane by N^{-1}

$$
\begin{aligned}
\mathsf{D}_\infty &= \mathbf{I}\mathbf{J}^\mathsf{T} + \mathbf{J}\mathbf{I}^\mathsf{T} \\
&= \mathsf{N}^{-1}\mathbf{I}'\mathbf{J}'^\mathsf{T}\mathsf{N}^{-\mathsf{T}} + \mathsf{N}^{-1}\mathbf{J}'\mathbf{I}'^\mathsf{T}\mathsf{N}^{-\mathsf{T}} \\
&= \begin{pmatrix} \alpha^2+\beta^2 & \alpha & \frac{-(\alpha^2+\beta^2)l_{\infty 1}-\alpha l_{\infty 2}}{l_{\infty 3}} \\ \alpha & 1 & \frac{-(\alpha l_{\infty 1}+l_{\infty 2})}{l_{\infty 3}} \\ \frac{-(\alpha^2+\beta^2)l_{\infty 1}-\alpha l_{\infty 2}}{l_{\infty 3}} & \frac{-(\alpha l_{\infty 1}+l_{\infty 2})}{l_{\infty 3}} & \frac{(\alpha l_{\infty 1}+l_{\infty 2})^2+\beta^2 l_{\infty 1}^2}{l_{\infty 3}^2} \end{pmatrix}
\end{aligned}
\tag{3.12}
$$

Once D_∞ is determined, the four rectification parameters are readily extracted from the matrix elements and property 3 below.

Properties of D_∞

1. **Incidence.** D_∞ is dual conic for which

$$\mathrm{l}^\mathsf{T}\mathsf{D}_\infty\mathrm{l} = \mathrm{l}^\mathsf{T}\mathbf{I}\mathbf{J}^\mathsf{T}\mathrm{l} + \mathrm{l}^\mathsf{T}\mathbf{J}\mathbf{I}^\mathsf{T}\mathrm{l} = 0$$

   for any line l incident with either $\mathbf{I}$ or $\mathbf{J}$, since then either $\mathrm{l}^\mathsf{T}\mathbf{I} = 0$ or $\mathrm{l}^\mathsf{T}\mathbf{J} = 0$.

2. **Rank.** A conic dual to a pair of points is degenerate, and its representative matrix of rank 2. In this particular case, note from (3.11) that D'_∞ is clearly rank 2. D_∞ remains rank 2 under any (full rank) projective transformation.

3. **Null vector.** The null vector of D_∞ is l_∞, the line incident with both $\mathbf{I}$ and $\mathbf{J}$:

$$\mathsf{D}_\infty\mathrm{l}_\infty = \mathbf{I}\mathbf{J}^\mathsf{T}\mathrm{l}_\infty + \mathbf{J}\mathbf{I}^\mathsf{T}\mathrm{l}_\infty = 0\mathbf{I} + 0\mathbf{J} = \mathbf{0}$$

4. **Polarity.** The pole of a line l with respect to D_∞ is a point

$$\mathbf{x} = \mathsf{D}_\infty\mathrm{l}$$

   The pole is incident with l_∞, since $\mathbf{x}^\mathsf{T}\mathrm{l}_\infty = \mathrm{l}^\mathsf{T}\mathsf{D}_\infty\mathrm{l}_\infty = 0$, and its position on l_∞ depends on

the direction of $\mathbf{l}$.

Consider a line $\mathbf{l}' = (l'_1,\ l'_2,\ l'_3)^\mathsf{T}$ on the metric plane:

$$\mathbf{x}' = \mathsf{D}'_\infty \mathbf{l}' = (l'_1,\ l'_2,\ 0)^\mathsf{T}$$

D'_∞ 'extracts' the directional information from the line $\mathbf{l}$. Additionally, the magnitude of the direction vector is given by

$$l_1'^{\,2} + l_2'^{\,2} = \mathbf{l}'^\mathsf{T} \mathsf{D}'_\infty \mathbf{l}'$$

5. **Angle.** Given a second line $\mathbf{m}'$, property 4 provides a compact expression for the inner product of the directional components of two lines $\mathbf{l}$ and $\mathbf{m}$:

$$\mathbf{m}'^\mathsf{T} \mathsf{D}'_\infty \mathbf{l}' = (m'_1,\ m'_2)(l'_1,\ l'_2)^\mathsf{T} = (\mathbf{m}'^\mathsf{T} \mathsf{D}'_\infty \mathbf{m}')^{\frac{1}{2}} (\mathbf{l}'^\mathsf{T} \mathsf{D}'_\infty \mathbf{l}')^{\frac{1}{2}} \cos\theta \tag{3.13}$$

The advantage of (3.13) is that it can be mapped under a projective transformation. The result

$$\cos(\theta) = \frac{\mathbf{m}^\mathsf{T} \mathsf{D}_\infty \mathbf{l}}{(\mathbf{m}^\mathsf{T} \mathsf{D}_\infty \mathbf{m})^{\frac{1}{2}} (\mathbf{l}^\mathsf{T} \mathsf{D}_\infty \mathbf{l})^{\frac{1}{2}}} \tag{3.14}$$

is a restatement of Laguerre's formula[91], and allows Euclidean measurement of angle to be made on the projective plane.

6. **Conjugacy.** When $\mathbf{m}$ and $\mathbf{l}$ are orthogonal $\cos\theta = 0$ and (3.14) simplifies to a *conjugacy.* A pair of *orthogonal* lines $\mathbf{l}$ and $\mathbf{m}$ are conjugate with respect to D_∞ and

$$\mathbf{m}^\mathsf{T} \mathsf{D}_\infty \mathbf{l} = 0 \tag{3.15}$$

Due to both the prevalence of orthogonalities in the man-made world and the non-linearity of (3.14), the investigations described in the following sections are mostly restricted to right angles. It will be shown in the next section how (3.15) is used to linearly constrain D_∞ from orthogonal scene lines.

3.5.2 Computing D_∞

D_∞ is a symmetric matrix with six elements. However, D_∞ is defined up to scale and is rank two, leaving five elements that must be computed and the rank constraint $\det(\mathsf{D}_\infty) = 0$. That leaves four degrees of freedom, agreeing with the parametrization of (3.12).

D_∞ can be computed from orthogonal lines on the image plane using (3.15). Each pair of orthogonal lines gives one *linear* equation in the six elements of D_∞

$$\begin{aligned}
\mathbf{m}^\mathsf{T}\mathsf{D}_\infty\mathbf{l} &= \begin{pmatrix} m_1 & m_2 & m_3 \end{pmatrix} \begin{pmatrix} \mathsf{d}_1 & \mathsf{d}_2 & \mathsf{d}_4 \\ \mathsf{d}_2 & \mathsf{d}_3 & \mathsf{d}_5 \\ \mathsf{d}_4 & \mathsf{d}_5 & \mathsf{d}_6 \end{pmatrix} \begin{pmatrix} l_1 \\ l_2 \\ l_3 \end{pmatrix} \\
&= m_1 l_1 \mathsf{d}_1 + (m_2 l_1 + m_1 l_2)\mathsf{d}_2 + m_2 l_2 \mathsf{d}_3 + (m_3 l_1 + m_1 l_3)\mathsf{d}_4 \\
&\quad + (m_3 l_2 + m_2 l_3)\mathsf{d}_5 + m_3 l_3 \mathsf{d}_6 \\
&= \boldsymbol{\varpi}^\mathsf{T}\mathbf{d} = 0
\end{aligned} \tag{3.16}$$

where $\mathbf{d} = (\mathsf{d}_1,\ \mathsf{d}_2,\ \ldots,\ \mathsf{d}_6)^\mathsf{T}$ and $\boldsymbol{\varpi}$ is the vector of coefficients.

A number of line pairs can be combined into a constraint matrix for which

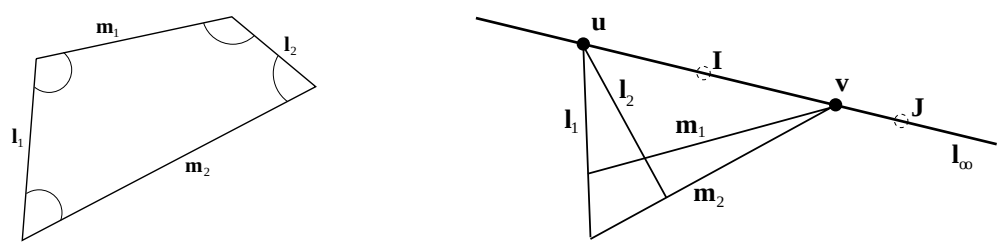
$$\mathsf{A}_{\boldsymbol{\varpi}}\mathsf{d} = \begin{pmatrix} \boldsymbol{\varpi}_1^\mathsf{T} \\ \vdots \\ \boldsymbol{\varpi}_n^\mathsf{T} \end{pmatrix} \mathsf{d} = \mathbf{0} \tag{3.17}$$

If five or more line pairs are available, D_∞ can then be computed as the null vector of $\mathsf{A}_{\boldsymbol{\varpi}}$. In the presence of noise, the vector associated with the smallest singular value of $\mathsf{A}_{\boldsymbol{\varpi}}$ is the estimate of d that minimizes $\|\mathsf{A}_{\boldsymbol{\varpi}}\mathsf{d}\|$ subject to the constraint $\|\mathsf{d}\| = 1$.

With only four orthogonal line pairs, the rank constraint must be applied. This is a cubic constraint with three solutions. Additionally, there is a degeneracy associated with the four constraints obtained from parallel lines in two dimensions - the constraints obtained from a rectangle. This is explored in the next section.

3.5.3 The rectangle ambiguity

The four angles at the vertices of a rectangle *do not* provide four constraints on D_∞. Geometrically it is obvious when one considers the nature of the four orthogonality constraints. Describing three of the four angles at the corners of a rectangle, as in figure 3.14 (a), as right angles implies that the fourth is also a right angle. These four constraints do not then constrain D_∞ entirely, but offer three of its four degrees of freedom. This is exactly the ambiguity described in section 3.4, where the orthogonality of two pairs of parallel lines determines the plane rectification up to an unknown aspect ratio. Algebraically, the rank of A_{ϖ} for four line pair constraints is four. It is significant that four independent constraints on the elements of D_∞ are not four independent equations in the four rectification parameters. It will be shown now that the reason for this is that the rank constraint must be satisfied by all members of the family of D_∞'s defined by the four orthogonalities of a rectangle.



(a) (b)

Figure 3.14: The four orthogonalities of a rectangle.

Using the notation of figure 3.14 (b), the two parallel side pairs of the rectangle are $\mathbf{l}_1, \mathbf{l}_2$ and $\mathbf{m}_1, \mathbf{m}_2$, with vanishing points $\mathbf{u}$ and $\mathbf{v}$. The four constraints defined by the orthogonality of line pairs are

$$\begin{aligned}
\mathbf{l}_1^\mathsf{T}\mathsf{D}_\infty\mathbf{m}_1 &= 0 \\
\mathbf{l}_1^\mathsf{T}\mathsf{D}_\infty\mathbf{m}_2 &= 0 \\
\mathbf{l}_2^\mathsf{T}\mathsf{D}_\infty\mathbf{m}_1 &= 0 \\
\mathbf{l}_2^\mathsf{T}\mathsf{D}_\infty\mathbf{m}_2 &= 0
\end{aligned} \tag{3.18}$$

Four constraints on five parameters implies a one parameter family of solutions of the form

$$\mathsf{D}_\infty(\mu,\nu) = \mu\mathsf{D}_{\infty 1} + \nu\mathsf{D}_{\infty_2} \tag{3.19}$$

for *homogeneous* parameter vector (μ,ν). It is possible to describe this family by defining the rank 1 degenerate conics dual to the single points

$$\mathsf{D}_{\infty 1} = \mathbf{u}\mathbf{u}^\mathsf{T} \text{ and } \mathsf{D}_{\infty 2} = \mathbf{v}\mathbf{v}^\mathsf{T} \tag{3.20}$$

It is simple to verify that $\mathsf{D}_{\infty 1}$ and $\mathsf{D}_{\infty 2}$ satisfy all of (3.18). In additional, they are both rank one, and thus individually and in any linear combination satisfy the rank constraint on D_∞. The rank constraint thus provides no further information and there remains a one parameter family of conics satisfying all available constraints.

Finally, note that the directions defined by $\mathbf{u}$ and $\mathbf{v}$ are orthogonal, so the four points on $\mathbf{l}_\infty$, $\mathbf{u}$, $\mathbf{v}$, $\mathbf{I}$ and $\mathbf{J}$, are *harmonic*: the cross ratio $\mathfrak{C}(\mathbf{u},\mathbf{v};\mathbf{I},\mathbf{J}) = -1$.

Resolving the Ambiguity.

An additional orthogonal line pair, at a different orientation to the rectangle sides, resolves the ambiguity. If lines $\mathbf{s}$ and $\mathbf{t}$ are orthogonal

$$\mathbf{s}^\mathsf{T}(\mu\mathsf{D}_{\infty 1} + \nu\mathsf{D}_{\infty 2})\mathbf{t} = \mathbf{s}^\mathsf{T}(\mu\mathbf{u}\mathbf{u}^\mathsf{T} + \nu\mathbf{v}\mathbf{v}^\mathsf{T})\mathbf{t} = 0$$

can be solved for μ/ν. In an alternative implementation, (3.16) can be used to add a fifth constraint to A_ϖ. An example of the diagonals of a square providing the extra constraint appears in figure 3.15.

The orthogonality of the diagonals of a square can also be interpreted as defining the ratio of lengths of the sides of the square. The constraint is thus related to the stratified constraints of section 3.3. Other angle and length ratio constraints are non-linear (as in (3.14)), but could also be used to constrain D_∞.

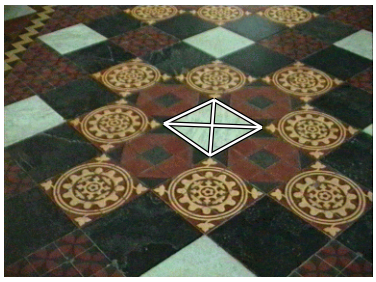


Figure 3.15: A square floor tile provides all the constraints necessary to compute the rectification parameters directly. The lines shown provide five orthogonality constraints, and the computed parameters are $\mathbf{l}_\infty = (0.0644,\ 0.9979,\ 442.0555)^\mathsf{T}$, $\alpha = -1.0108$ and $\beta = 2.9108$. Note that α and β are different from the example in figure 3.11 because in that case the affine rectification was rotated before affine constraints were computed.

3.6 Maximum Likelihood line intersection

The intersection of parallel lines to compute a vanishing point is an important feature of the plane rectification techniques presented in this chapter. Vanishing points continue to play an extremely important role in the calibration and reconstruction methods of the following chapters. In addition, line intersection is commonly used to define feature points in an image. It is thus crucial to compute an optimal estimate for the intersection of a number of lines. The problem is that, due to noise in measurement, lines which should be coincident do not in fact intersect in a unique point. This section describes a Maximum Likelihood Estimate (MLE) of the intersection point and the error in the computed point.

3.6.1 ML estimation

A number of approaches to estimating the vanishing point have been proposed. Caprile and Torre [11] calculate of a weighted mean of all pairwise line intersections. A more elaborate technique by Collins involves the application of Bayesian statistics to error in projective spaces [16]. Kanatani [56] computes vanishing points from the null space of a set of normalised lines. A number of methods also exist to compute vanishing points automatically, for example using Hough transform methods [15, 94, 22]. The estimation and errors are then dependent on the Hough parameters. The

approach taken here assumes that the set of coincident lines is known. The problem is then to compute an estimate of the point of intersection of a number of *given* line segments in a way that minimizes a meaningful error.

The error occurs in the image, in the line segments computed from image gradient techniques such as Canny edge detection[10], and in the subsequent processing: edge linking; segmentation of the edgel chain at high curvature points; and finally, straight line fitting by orthogonal regression to the resulting chain segments. It is thus sensible to minimize the errors where they occur. It is assumed below that the error in the fitted line segments can be modeled by isotropic zero mean Gaussian noise on the endpoints. However, this is done for demonstration purposes and simplifies the analysis; it does not preclude the development of a more accurate model of the uncertainty which could, for example, start with image noise and propagate the subsequent uncertainty through the various algorithm steps.

The MLE is defined by considering the *likelihood*: a probability density function (pdf) describing the probability of obtaining a measurement given the true value of a variable[49]. In this case, it is assumed that the likelihood for each line segment endpoint has a zero mean isotropic Gaussian distribution, and that each of the endpoint measurements are independent and have the same variance σ^2. Consequently, the likelihood function for the set of all line segment endpoints is the product of the Gaussian pdfs of each of the measurements. Explicitly, for the set of measured line segments, with endpoints $\mathbf{a}_i$ and $\mathbf{b}_i$ which have true values $\overline{\mathbf{a}}_i$ and $\overline{\mathbf{b}}_i$ defining concurrent line segments, the likelihood is

$$P(\{\mathbf{a}_i, \mathbf{b}_i\}|\{\overline{\mathbf{a}}_i, \overline{\mathbf{b}}_i\}) = \prod_i \frac{1}{2\pi\sigma^2} e^{-\frac{d^2(\mathbf{a}_i, \overline{\mathbf{a}}_i) + d^2(\mathbf{b}_i, \overline{\mathbf{b}}_i)}{2\sigma^2}}$$

where $d^2(\mathbf{a}, \overline{\mathbf{a}}_i)$ is the squared distance between points $\mathbf{a}$ and $\overline{\mathbf{a}}_i$.

The MLE is then the estimate of endpoints and intersection point that maximises the *log-likelihood* function, and thus minimises the geometric cost function

$$\mathcal{C} = \sum_i d^2(\hat{\mathbf{a}}_i, \mathbf{a}_i) + d^2(\hat{\mathbf{b}}_i, \mathbf{b}_i) = \sum_i d_\perp^2(\mathbf{l}_i, \mathbf{a}_i) + d_\perp^2(\mathbf{l}_i, \mathbf{b}_i)$$

subject to the constraints $\mathbf{u}^\mathsf{T}\mathbf{l}_i = 0 \; \forall \, i$. The points $\hat{\mathbf{a}}_i$ and $\hat{\mathbf{b}}_i$ are estimates of the true values, and define a set of line segments $\mathbf{l}_i$ which intersect in point $\mathbf{u}$ (figure 3.6.1). An alternative cost function under different noise assumptions is given by Kanatani [57].

In fact, the cost function $\mathcal{C}$ is the Mahalanobis distance between the vector of inhomogeneous measured endpoints $\mathbf{V}_{\tilde{\mathbf{a}},\tilde{\mathbf{b}}}$ and the vector of inhomogeneous endpoint estimates $\mathbf{V}_{\bar{\tilde{\mathbf{a}}},\bar{\tilde{\mathbf{b}}}}$

$$\|\mathbf{V}_{\tilde{\mathbf{a}},\tilde{\mathbf{b}}} - \mathbf{V}_{\bar{\tilde{\mathbf{a}}},\bar{\tilde{\mathbf{b}}}}\|^2_\Lambda = (\mathbf{V}_{\tilde{\mathbf{a}},\tilde{\mathbf{b}}} - \mathbf{V}_{\bar{\tilde{\mathbf{a}}},\bar{\tilde{\mathbf{b}}}})^\mathsf{T}\Lambda^{-1}(\mathbf{V}_{\tilde{\mathbf{a}},\tilde{\mathbf{b}}} - \mathbf{V}_{\bar{\tilde{\mathbf{a}}},\bar{\tilde{\mathbf{b}}}})$$

where Λ is the covariance matrix of the inhomogeneous vector of endpoint segments. Since variance is equal for all endpoints, the relative weighting induced by Λ falls out of the cost. It could, however, be reintroduced given a more accurate model of the endpoint noise.

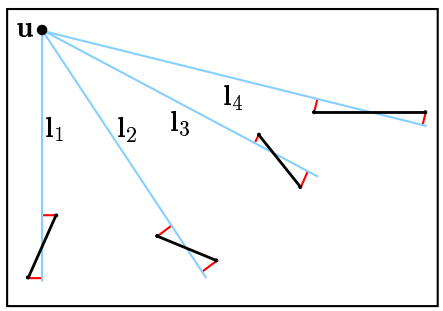


Figure 3.16: MLE vanishing point estimation: The vanishing point $\mathbf{u}$ is estimated as the intersection of the fitted lines $\mathbf{l}_i$ (in gray), which minimize the orthogonal distances from the endpoints of measured imaged parallel line segments (shown in black).

The method of finding the intersection point is then to apply a numerical minimization, varying the position of $\mathbf{u}$ and computing $\mathcal{C}$. However, for each $\mathbf{u}$ there is a pencil of lines for which $d^2_\perp$ could be computed. It is therefore necessary to compute the *minimum* error for each line segment given the point $\mathbf{u}$.

The minimum error given $\mathbf{u}$

Objective: Given $\mathbf{a}$ and $\mathbf{b}$, the endpoints of a line segment, and a third point $\mathbf{u}$, find the error associated with the line through $\mathbf{u}$ minimizing the perpendicular error to $\mathbf{a}$ and $\mathbf{b}$, as shown in figure

3.17.

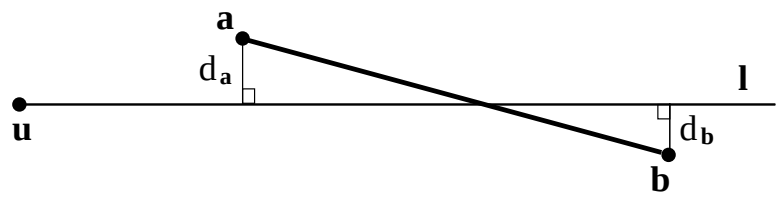


Figure 3.17: Geometry of the cost function: given point $\mathbf{u}$ and the line segment with endpoints $\mathbf{a}$ and $\mathbf{b}$, error is measured by $d_{\mathbf{a}}^2 + d_{\mathbf{b}}^2$

Since the endpoints of the line are observations in an image, they maybe assumed to be finite and written as $\mathbf{a} = (a_1,\ a_2,\ 1)^\mathsf{T}$ and $\mathbf{b} = (b_1,\ b_2,\ 1)^\mathsf{T}$. Point $\mathbf{u} = (u_1,\ u_2,\ u_3)^\mathsf{T}$ is treated homogeneously. The error is written with respect to the line $\mathrm{l} = (l_1,\ l_2,\ l_3)^\mathsf{T}$, and the cost function is then

$$\mathcal{F}(\mathrm{l}) = d_{\mathbf{a}}^2 + d_{\mathbf{b}}^2 = \frac{(\mathbf{a}^\mathsf{T}\mathrm{l})^2 + (\mathbf{b}^\mathsf{T}\mathrm{l})^2}{l_1^2 + l_2^2} = \frac{(\mathbf{a}^\mathsf{T}\mathrm{l})^2 + (\mathbf{b}^\mathsf{T}\mathrm{l})^2}{\mathrm{l}^\mathsf{T}\mathsf{D}'_\infty\mathrm{l}} \tag{3.21}$$

Now, l can be parametrized by two points. Since $\mathbf{u}$ lies on l and l must also intersect line $\mathbf{ab}$, choose the points $\mathbf{u}$ and $\mathbf{c} = (k_{\mathbf{a}}\mathbf{a} + k_{\mathbf{b}}\mathbf{b})$. Then

$$\begin{aligned} \mathrm{l} &= \mathbf{u} \times (k_{\mathbf{a}}\mathbf{a} + k_{\mathbf{b}}\mathbf{b}) \\ &= \mathbf{u} \times [\mathbf{a}\ \mathbf{b}]\,(k_{\mathbf{a}}\ k_{\mathbf{b}})^\mathsf{T} \\ &= [\mathbf{u}]_\times\,[\mathbf{a}\ \mathbf{b}]\,\mathbf{k} \end{aligned} \tag{3.22}$$

and $\mathcal{F}$ is

$$\begin{aligned}
\mathcal{F}(\mathbf{k}) &= \frac{(\mathbf{a}^\top(\mathbf{u}\times[\mathbf{a}\ \mathbf{b}]\,\mathbf{k}))^2 + (\mathbf{b}^\top(\mathbf{u}\times[\mathbf{a}\ \mathbf{b}]\,\mathbf{k}))^2}{([\mathbf{u}]_\times\,[\mathbf{a}\ \mathbf{b}]\,\mathbf{k})^\top \mathsf{D}'_\infty([\mathbf{u}]_\times\,[\mathbf{a}\ \mathbf{b}]\,\mathbf{k})} \\
&= \frac{(\mathbf{u}^\top(\mathbf{a}\times[\mathbf{a}\ \mathbf{b}]\,\mathbf{k}))^2 + (\mathbf{u}^\top(\mathbf{b}\times[\mathbf{a}\ \mathbf{b}]\,\mathbf{k}))^2}{\mathbf{k}^\top[\mathbf{u}]_\times\,[\mathbf{a}\ \mathbf{b}]^\top\,\mathsf{D}'_\infty[\mathbf{u}]_\times\,[\mathbf{a}\ \mathbf{b}]\,\mathbf{k}} \\
&= \frac{(\mathbf{u}^\top([\mathbf{0}\ (\mathbf{a}\times\mathbf{b})]\,\mathbf{k}))^2 + (\mathbf{u}^\top([(\mathbf{a}\ \times\mathbf{b})\ \mathbf{0}]\,\mathbf{k}))^2}{\mathbf{k}^\top[\mathbf{u}]_\times\,[\mathbf{a}\ \mathbf{b}]^\top\,\mathsf{D}'_\infty[\mathbf{u}]_\times\,[\mathbf{a}\ \mathbf{b}]\,\mathbf{k}} \\
&= \frac{2(\mathbf{u}^\top(\mathbf{a}\times\mathbf{b}))^2\mathbf{k}^\top\mathbf{k}}{\mathbf{k}^\top\,[\mathbf{a}\ \mathbf{b}]^\top\,[\mathbf{u}]_\times^\top\mathsf{D}'_\infty[\mathbf{u}]_\times\,[\mathbf{a}\ \mathbf{b}]\,\mathbf{k}} \\
&= \frac{\mathbf{k}^\top\mathbf{k}}{\mathbf{k}^\top\mathsf{A}\mathbf{k}}
\end{aligned} \tag{3.23}$$

where

$$\mathsf{A} = \frac{1}{2(\mathbf{u}^\top(\mathbf{a}\times\mathbf{b}))^2}\,[\mathbf{a}\ \mathbf{b}]^\top\,[\mathbf{u}]_\times^\top\mathsf{D}'_\infty[\mathbf{u}]_\times\,[\mathbf{a}\ \mathbf{b}] = \frac{1}{\mu}\begin{bmatrix} A_{11} & A_{12} \\ A_{12} & A_{22} \end{bmatrix}$$

$$\begin{aligned}
\mu &= 2(b_2u_1 - b_1u_2 - u_1a_2 + b_1u_3a_2 + u_2a_1 - b_2u_3a_1)^2 \\
A_{11} &= (-a_2u_3 + u_2)^2 + (-u_1 + u_3a_1)^2 \\
A_{12} &= (-a_2u_3 + u_2)(u_2 - b_2u_3) + (-u_1 + u_3a_1)(b_1u_3 - u_1) \\
A_{22} &= (u_2 - b_2u_3)^2 + (b_1u_3 - u_1)^2
\end{aligned}$$

Following the Courant-Fischer minimax theorem [39], the minimum of the cost function $\mathcal{F}(\mathbf{k})$ is given by the inverse of the larger of the eigenvalues of A. Taking the larger of the roots of the characteristic polynomial of A:

$$\mathcal{F}_{\min} = \frac{\mu}{A_{11} + A_{22} + \sqrt{(A_{11} - A_{22})^2 + 4A_{12}^2}}$$

The qualitative behaviour of the fitted line $\mathbf{l}$ is shown in figure 3.18, where the point $\mathbf{u}$ is moved relative to the line segment endpoints and its intersection with the measured line segment illustrated. With $\mathbf{u}$ close to the line and equidistant from $\mathbf{a}$ and $\mathbf{b}$, $\mathbf{l}$ is parallel to the line segment, eventually

intersecting the line segment at its midpoint when $\mathbf{u}$ is ideal.

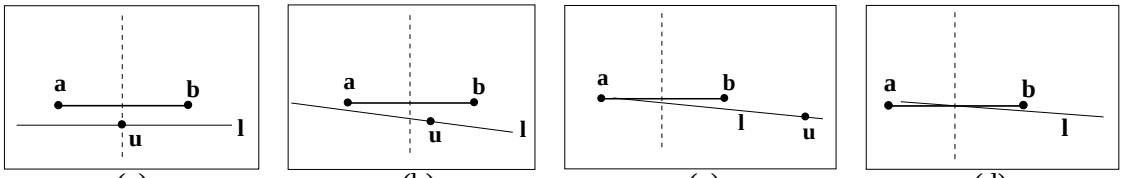


Figure 3.18: The intersection of the MLE line with the measured line segment. (a) An extreme example, with $\mathbf{u}$ equidistant from $\mathbf{a}$ and $\mathbf{b}$. Fitted line $\mathbf{l}$ is parallel to the measured line segment. (b) and (c) As $\mathbf{u}$ moves further from the endpoints, $\mathbf{l}$ approaches the mid point of the line segment. (d) When $\mathbf{u}$ is *ideal*, $\mathbf{l}$ intersects the measured line segment mid way between $\mathbf{a}$ and $\mathbf{b}$.

Computing $\mathbf{u}$

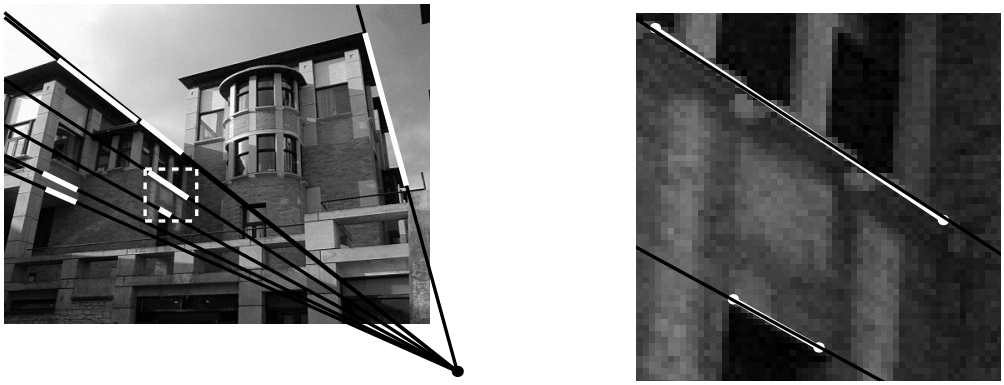
Once $\mathcal{F}_{\text{min}}$ can be computed, $\mathcal{C}(\mathbf{u})$ can be minimized over $\mathbf{u}$ using the Levenberg-Marquardt numerical algorithm [82] to minimize

$$\mathcal{C} = \sum_i \mathcal{F}_{\text{min}}^{(i)} = \sum_i \frac{\mu^{(i)}}{A_{11}^{(i)} + A_{22}^{(i)} + \sqrt{(A_{11}^{(i)} - A_{22}^{(i)})^2 + 4A_{12}^{2(i)}}} \tag{3.24}$$

In practice, least square algorithms of this sort are implemented to compute a vector of residual error terms. It is therefore necessary to compute the eigenvectors of A in (3.23), $\mathbf{l}_i$ from (3.22) and thereby $d_{\mathbf{a}}$ and $d_{\mathbf{b}}$. An initial solution for $\mathbf{u}$ is obtained from the null vector of the matrix $(\mathbf{l}_1, \mathbf{l}_2, \ldots \mathbf{l}_n)$ via singular value decomposition (SVD). Experience indicates that with suitable normalisation the SVD provides a very good initial estimate of the point of intersection, since it is the point which minimizes an algebraic error. An example MLE vanishing point appears in figure 3.19.

3.6.2 Error propagation

Given the assumption of normally distributed error on the endpoints of the parallel line segments, it is natural to ask what the error in the computed point of intersection is. Furthermore, it is useful to know how the error distribution changes with data. It would be expected, for example, that increasing the number of parallel segments would reduce the uncertainty in the intersection point, and that lines far apart in the image would result in a different covariance profile from those close



(a) (b)

Figure 3.19: Seven parallel lines. (a) The fitted lines (black) minimize distance to the endpoints of the measured parallel line segments (in white) and intersect in the MLE of the vanishing point shown. Note that parallel lines do not need to be co-planar. (b) A close-up of two of the line segments, showing the small (but noticeable) deviation between the fitted lines and the image line segments.

together. To answer these questions, this section presents a first order error propagation. Given the covariance matrix of the image line segment endpoints, an estimate of the covariance of the intersection point is computed. The section proceeds by first computing the covariance of a point obtained from the intersection of two lines using the Jacobian of the cross product. This is followed by the covariance of the ML estimate. In the latter case, the numerical estimate of the Jacobian computed as part of the iterative gradient descent algorithm is used. Monte Carlo simulations verify the use of first order error models, and the section concludes with a discussion of the behaviour of the error under various conditions.

Error with two line segments

Given only two line segments, the point of intersection $\mathbf{u}$ is, writing $\mathbf{a}_i = (a_{i1},\ a_{i2},\ 1)^\mathsf{T}$ and $\mathbf{b}_i = (b_{i1},\ b_{i2},\ 1)^\mathsf{T}$,

$$\begin{aligned}\mathbf{u} &= \mathbf{l}_1 \times \mathbf{l}_2 = (\mathbf{a}_1 \times \mathbf{b}_1) \times (\mathbf{a}_2 \times \mathbf{b}_2) \\ &= \begin{pmatrix} (b_{11}-a_{11})(a_{21}b_{22}-a_{22}b_{21}) - (a_{11}b_{12}-a_{12}b_{11})(b_{21}-a_{21}) \\ (a_{11}b_{12}-a_{12}b_{11})(a_{22}-b_{22}) - (a_{12}-b_{12})(a_{21}b_{22}-a_{22}b_{21}) \\ (a_{12}-b_{12})(b_{21}-a_{21}) - (b_{11}-a_{11})(a_{22}-b_{22}) \end{pmatrix}\end{aligned}$$

The line segment endpoints can be shaped into a vector

$$\mathbf{e}_2 = \begin{pmatrix} a_{11} & a_{12} & b_{11} & b_{12} & a_{21} & a_{22} & b_{21} & b_{22} \end{pmatrix}^\mathsf{T}$$

The (homogeneous) covariance of $\mathbf{u}$, $\Lambda_\mathbf{u}$, can be estimated from the Jacobian $\mathsf{J}_\mathbf{u}$ of the function mapping the line segment endpoints to $\mathbf{u}$[31, 13]

$$\mathsf{J}_\mathbf{u}^\mathsf{T} = \nabla\mathbf{u} = \begin{pmatrix} \frac{\partial u_1}{\partial \mathbf{e}_2}^\mathsf{T} & \frac{\partial u_2}{\partial \mathbf{e}_2}^\mathsf{T} & \frac{\partial u_3}{\partial \mathbf{e}_2}^\mathsf{T} \end{pmatrix} \tag{3.25}$$

and the covariance of the endpoint vector $\Lambda_{\mathbf{e}_2}$:

$$\Lambda_\mathbf{u} = \mathsf{J}_\mathbf{u}\Lambda_{\mathbf{e}_2}\mathsf{J}_\mathbf{u}^\mathsf{T}$$

The Jacobian has components

$$\frac{\partial u_1^{\mathsf{T}}}{\partial \mathbf{e}_2} = \begin{pmatrix} -a_{21}b_{22} + a_{22}b_{21} - b_{12}(b_{21} - a_{21}) \\ b_{11}(b_{21} - a_{21}) \\ a_{21}b_{22} - a_{22}b_{21} + a_{12}(b_{21} - a_{21}) \\ -a_{11}(b_{21} - a_{21}) \\ (b_{11} - a_{11})b_{22} + a_{11}b_{12} - a_{12}b_{11} \\ -(b_{11} - a_{11})b_{21} \\ -(b_{11} - a_{11})a_{22} - a_{11}b_{12} + a_{12}b_{11} \\ (b_{11} - a_{11})a_{21} \end{pmatrix},$$

$$\frac{\partial u_2^{\mathsf{T}}}{\partial \mathbf{e}_2} = \begin{pmatrix} b_{12}(a_{22} - b_{22}) \\ -b_{11}(a_{22} - b_{22}) - a_{21}b_{22} + a_{22}b_{21} \\ -a_{12}(a_{22} - b_{22}) \\ a_{11}(a_{22} - b_{22}) + a_{21}b_{22} - a_{22}b_{21} \\ -(a_{12} - b_{12})b_{22} \\ a_{11}b_{12} - a_{12}b_{11} + (a_{12} - b_{12})b_{21} \\ (a_{12} - b_{12})a_{22} \\ -a_{11}b_{12} + a_{12}b_{11} - (a_{12} - b_{12})a_{21} \end{pmatrix},$$

and

$$\frac{\partial u_3^{\mathsf{T}}}{\partial \mathbf{e}_2} = \begin{pmatrix} a_{22} - b_{22} \\ b_{21} - a_{21} \\ -a_{22} + b_{22} \\ -b_{21} + a_{21} \\ -a_{12} + b_{12} \\ -b_{11} + a_{11} \\ a_{12} - b_{12} \\ b_{11} - a_{11} \end{pmatrix}$$

Inhomogeneous covariance

The point $\mathbf{u}$ is homogeneous, as is its covariance. The covariance matrix is defined up to scale and has rank two. To obtain an inhomogeneous point $\tilde{\mathbf{u}}$ and associated covariance, an additional transformation must be applied to $\mathbf{u}$. If $\mathbf{u}$ is finite the inhomogeneous point $\tilde{\mathbf{u}}$ is given by

$$\tilde{\mathbf{u}} = (\tilde{u}_1,\ \tilde{u}_2)^\mathsf{T} = (\frac{u_1}{u_3},\ \frac{u_2}{u_3})^\mathsf{T}$$

with Jacobian

$$\mathsf{J}_{\tilde{\mathbf{u}}} = \begin{pmatrix} \frac{1}{u_3} & 0 & \frac{-u_1}{u_3^2} \\ 0 & \frac{1}{u_3} & \frac{-u_2}{u_3^2} \end{pmatrix}$$

and the covariance of the inhomogeneous point follows as before:

$$\Lambda_{\tilde{\mathbf{u}}} = \mathsf{J}_{\tilde{\mathbf{u}}} \Lambda_{\mathbf{u}} \mathsf{J}_{\tilde{\mathbf{u}}}^\mathsf{T}$$

If the point is ideal, with $u_3 = 0$, the covariance of the direction vector $(u_1,\ u_2)^\mathsf{T}$ is the top left sub-matrix of $\Lambda_{\mathbf{u}}$.

Note that computations are performed on conditioned image co-ordinates, where the image bounds are scaled and translated to approximately the unit square centred on the origin. Conditioning simply scales the endpoint covariance matrix. The Jacobian of the deconditioning function is just the deconditioning matrix, and so poses no problem.

Error in the MLE

The first order estimate of the covariance of an MLE is a far cry from the straightforward case of an explicit function, such as the intersection of two lines. The method described below follows the description given by Hartley and Zisserman[49].

The two line segment case required the Jacobian of the function mapping $\mathbf{e}$ to $\mathbf{u}$, $\frac{\partial \mathbf{u}}{\partial \mathbf{e}}$. In the MLE case, however, the covariance of estimated intersection point $\mathbf{u}$ depends on the Jacobian of the function $\mathbf{f}$ that maps $\mathbf{u}$ to the projection of the measured data onto a space where it is *consistent*

with the estimate. That is, the vector of measured line segment endpoints

$$\mathbf{e} = (\mathbf{a}_{11},\ \mathbf{a}_{12},\ \mathbf{b}_{11},\ \mathbf{b}_{12},\ \ldots\ ,\ \mathbf{a}_{n1},\ \mathbf{a}_{n2},\ \mathbf{b}_{n1},\ \mathbf{b}_{n2})^{\mathsf{T}}$$

determines an ML estimate of intersection point $\mathbf{u}$. This point is mapped by $\mathbf{f}$ to the vector of line segment endpoints

$$\hat{\mathbf{e}} = (\hat{\mathbf{a}}_{11},\ \hat{\mathbf{a}}_{12},\ \hat{\mathbf{b}}_{11},\ \hat{\mathbf{b}}_{12},\ \ldots\ ,\ \hat{\mathbf{a}}_{n1},\ \hat{\mathbf{a}}_{n2},\ \hat{\mathbf{b}}_{n1},\ \hat{\mathbf{b}}_{n2})^{\mathsf{T}} = \mathbf{f}(\mathbf{u})$$

that define a set of lines concurrent with $\mathbf{u}$.

Hartley and Zisserman show that, under appropriate conditions, the covariance of $\mathbf{u}$ is then[2]

$$\Lambda_{\mathbf{u}} = (\mathsf{J}_f^{\mathsf{T}} \Lambda_{\mathbf{e}}^{-1} \mathsf{J}_f)^{+} \tag{3.26}$$

where J_f is the Jacobian of $\mathbf{f}$ given by

$$\mathsf{J}_f^{\mathsf{T}} = \nabla \hat{\mathbf{e}} = \frac{\partial \hat{\mathbf{e}}}{\partial \mathbf{u}}$$

evaluated at the estimated $\mathbf{u}$.

J_f has been determined analytically, but it is a large and cumbersome expression, much too large to be written down. It is also unnecessary, since the Levenberg-Marquardt function minimizes $\|\hat{\mathbf{e}} - \mathbf{e}\|$ it computes the gradient $\frac{\partial(\hat{\mathbf{e}}-\mathbf{e})}{\partial \mathbf{u}}$ as part of the minimization procedure. It is no surprise that it computes the gradient extremely accurately, with analytical and numerical estimates identical (up to scale) up to five significant figures in experiment. In fact, this applies even to the two line case, where the residual error is always zero, but the minimization algorithm computes the gradient in the vicinity of the solution as it verifies this.

[2]The notation A^{+} represents the pseudo-inverse of A.

Experiments

Armed with $\Lambda_{\mathbf{u}}$, the first issue to consider is how realistic the first order estimate is, and for what degree of noise. It would further be anticipated that the covariance changes as a function of the number of line segments used and the distance between them. These issues are investigated below. The experiments utilise synthetic data to which noise is added. To ensure a realistic data set, synthetic data is extracted from the example of figure 3.19 by using the closest endpoints to the measured segments on the *fitted lines*. The synthetic line segments are shown numbered in figure 3.20.

Figure 3.20: A set of synthetic parallel line segments extracted from a real example. The synthetic segments are obtained from the measured line segments defining edges of the structure (segment 7 is the result of merging three collinear roof edge segments) and using the fitted lines resulting from the MLE algorithm.

To begin, figures 3.21, 3.22, 3.23 and 3.24 show Monte Carlo simulations of line intersections for two cases. The first of these, figure 3.21, shows 10000 points computed as the intersection of two lines segments (segments 1 and 4 from figure 3.20). Mean zero isotropic Gaussian noise is added to the endpoints of the two line segments, the intersection point is computed and the covariance of the computed points calculated. The three standard deviation ellipse for the data with noise with a standard deviation of $\sigma = 1$ pixel is shown in figure 3.20 (a). This ellipse together with theoretical covariance ellipse, computed from (3.25), is plotted in (b). The same experiment with an additive noise $\sigma = 0.5$ pixels appears in figure 3.20 (c) and (d). Observe that the experimental and theoretical

ellipses are nearly indistinguishable in the figures. This indicates that the first order approximation is valid. The theoretical covariance ellipses for the two noise conditions are plotted on the same axes in figure 3.22, where the effect of noise magnitude is clearly visible. Note also the orientation of the covariance ellipses, the major axis approximately aligned with the directions of the intersecting lines.

Figure 3.23 is identical to the above experiment, but using all seven lines from figure 3.20. It is clear that the covariance is slightly underestimated by the first order covariance estimate for noise with $\sigma = 1$ pixel, although the difference is not evident when $\sigma = 0.5$ pixels.

The Monte Carlo comparisons suggest that the first order approximation to the covariance is a good one for the purposes at hand. Consider the noise levels used. A one pixel standard deviation implies the belief that there is a 99.7% chance of the true point lying within three pixels of the measured point. This kind of accuracy would be expected of a user selecting a clear feature point in an image with a mouse, while sub-pixel accuracies are often quoted for the more common methods of computing points, such as from correlation[6]. The estimates are certainly good enough to provide a qualitative sense of how the intersection point behaves. The effect of the number of line segments and the spacing between them are investigated in the following experiments.

Turning first to the number of line segments used: figure 3.25 plots the covariance ellipses for the intersection point of different numbers of line segments with $\sigma = 0.5$ pixels (the numbers in braces refers to figure 3.20). There is a consistent reduction in covariance as the number of line segments increases, although it is most marked for the two largest ellipses, both for two line segments (the larger of which corresponds to the segments used in Monte Carlo simulation). The larger covariance is computed using segments 1 and 4, which are closer together in the image compared to segments 1 and 7, and have a smaller angle between them. This suggests the next experiment, observing the effect of the angle between line segments on the uncertainty in the intersection point.

The effect of the relative orientation of line segments on the magnitude and orientation of covariance is investigated for the configuration shown in figure 3.26. A point of intersection $\mathbf{u}$ is fixed and two line segments with an angle θ to the horizontal are used to compute the covariance of $\mathbf{u}$ assuming endpoint noise of $\sigma = 0.5$ pixels. The lines are symmetric about the horizontal, so the

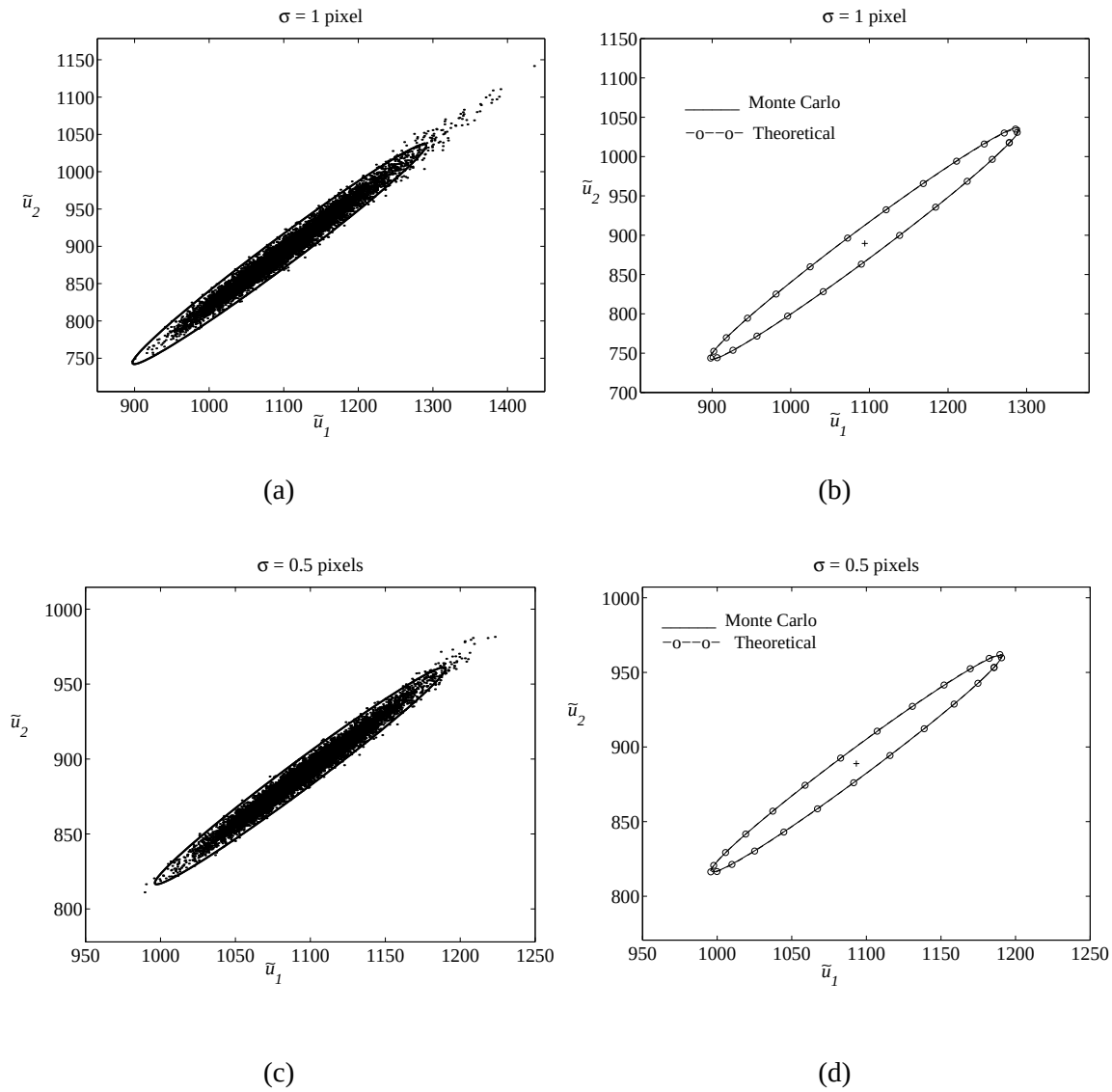


Figure 3.21: Monte Carlo simulation of the intersection of line segments 1 and 4 with isotropic Gaussian noise on the endpoints (a) σ = 1 pixel: 10000 points and the 3σ covariance ellipse. (b) σ = 1 pixel: The 10000 point covariance ellipse and the theoretical covariance from (3.25). The ellipses are virtually identical, with no difference visible at this scale. (c) σ = 0.5 pixels: 10000 points and the 3σ covariance ellipse. (d) σ = 0.5 pixels: The 10000 point covariance ellipse and the theoretical covariance. Again, the difference is not visible. As expected, the first order approximation is a good one, even at noise levels with a standard deviation of a pixel, and reduces with smaller noise levels (see figure 3.22).

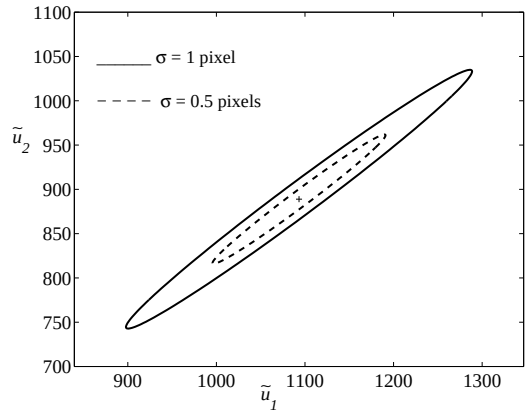


Figure 3.22: The (theoretical) covariance ellipses for the intersection of line segments 1 and 4 on the same axes. The effect of lower noise on the error in the estimated point is clear.

covariance ellipses are aligned with the axes.

Figure 3.27 shows the covariance ellipses obtained as the angle between lines is increased, and plots the standard deviations in $\tilde{u}_1$ and $\tilde{u}_2$. Note that at $\theta = 45^o$, when the lines are at right angles, the covariance plot is a circle and the error is isotropically distributed. The standard deviations as a function of θ indicate the the horizontal variance is high for lines with a small difference in orientation, and decreases rapidly as the angle increases. Vertical variance displays the inverse behaviour, increasing as the angle between lines increases.

3.7 Summary

Metric rectification of an imaged plane from scene geometry has been described in detail in this chapter, extending earlier work on affine rectification from the vanishing line of an imaged plane. The new techniques use geometric properties such as angle and relative length to compute the imaged circular points. This is done in both stratified and unstratified contexts.

The stratified approach first computes the vanishing line of the imaged plane, typically from parallel lines, and then defines quadratic constraints on the circular points in the affine rectified plane. These quadratic constraints take the form of circles on the plane of the two degrees of freedom of the circular points, the real and imaginary parts of the 1D coordinates of the circular

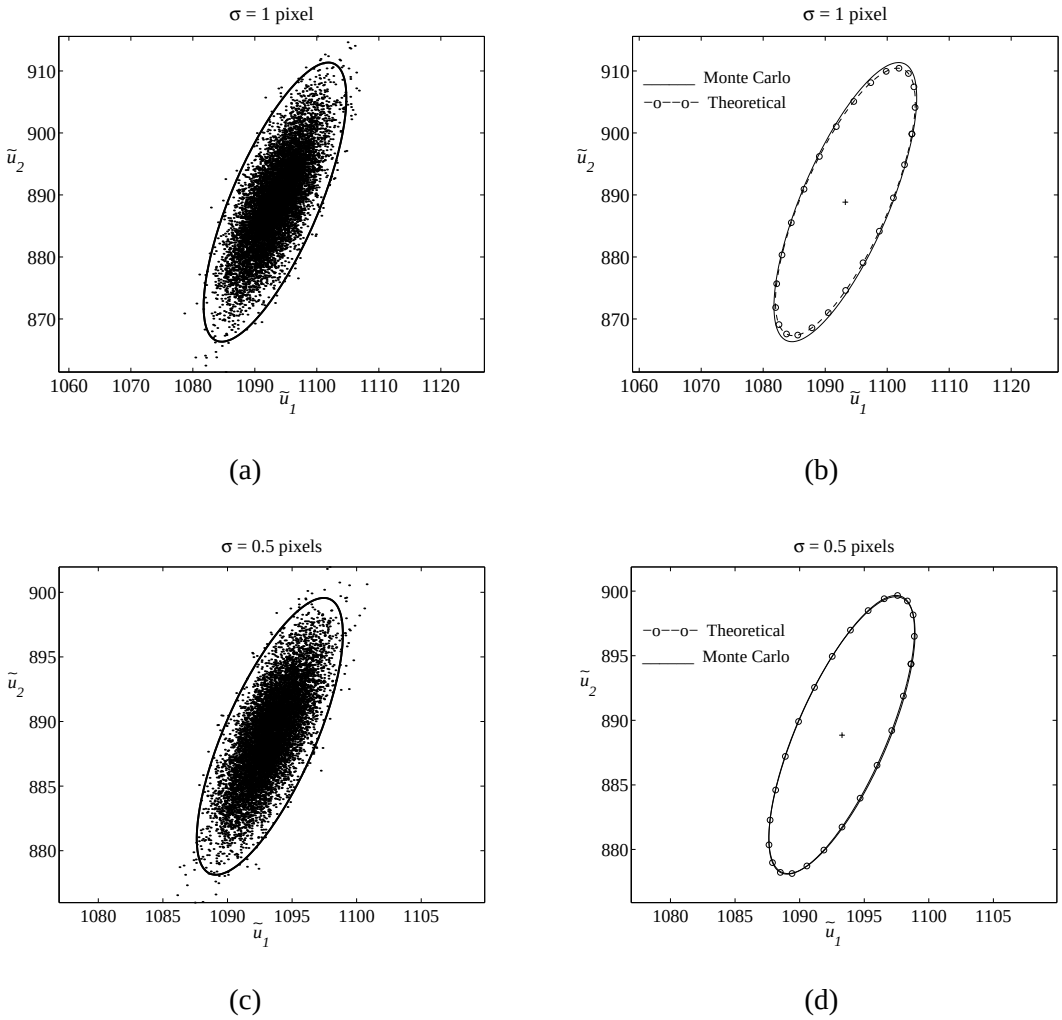


Figure 3.23: Monte Carlo simulation of the intersection of all seven line segments with isotropic Gaussian noise on the endpoints (a) $\sigma = 1$ pixel: 10000 points and the 3σ covariance ellipse. (b) $\sigma = 1$ pixel: The 10000 point covariance ellipse and the theoretical covariance from (3.25). The ellipses are close, but a small difference is visible at this scale. (c) $\sigma = 0.5$ pixels: 10000 points and the 3σ covariance ellipse. (d) $\sigma = 0.5$ pixels: The 10000 point covariance ellipse and the theoretical covariance. The difference is smaller, barely visible in this case. Again, the first order approximation is a good one, even at noise levels with a standard deviation of a pixel, and reduces with smaller noise levels (see figure 3.24). The relative errors as a function of the *number* of lines is compared in figure 3.25.

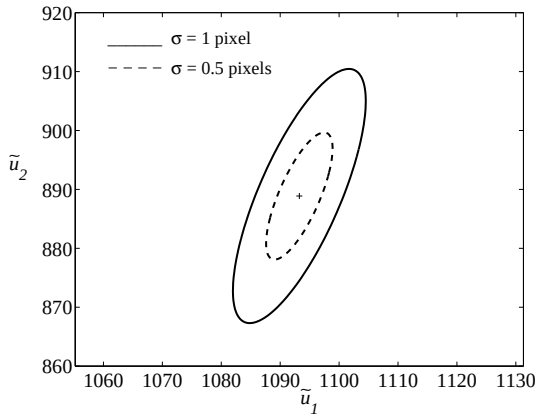
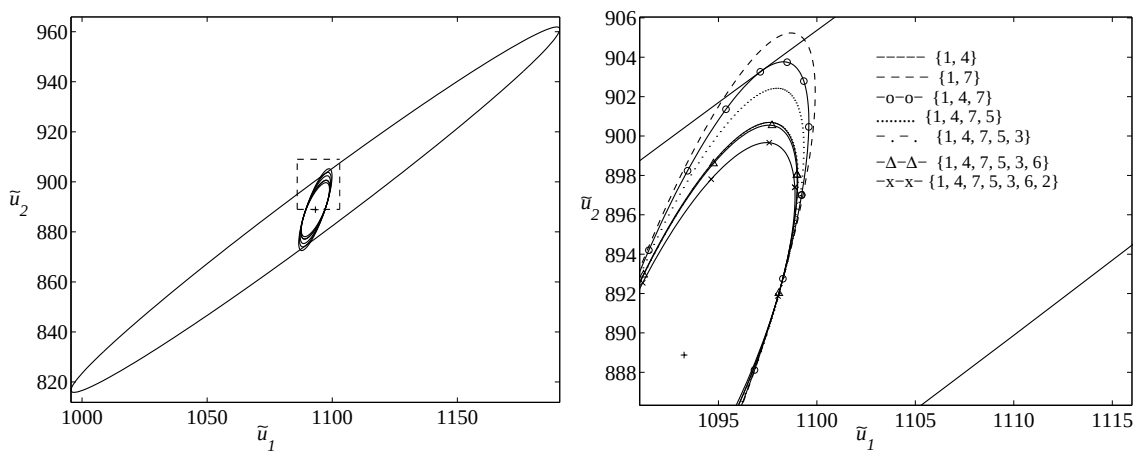


Figure 3.24: The (theoretical) covariance ellipses for the intersection of all seven line segments on n the same axes. The effect of lower noise on the error in the estimated point is clear.



(a) (b)

Figure 3.25: Covariance of an intersection point as a function of the number of line segments. 3σ covariance ellipses are drawn for different numbers of line segments, the numbers referring to figure 3.20. The covariance ellipses decrease in size with an increase in the number of line segments used.

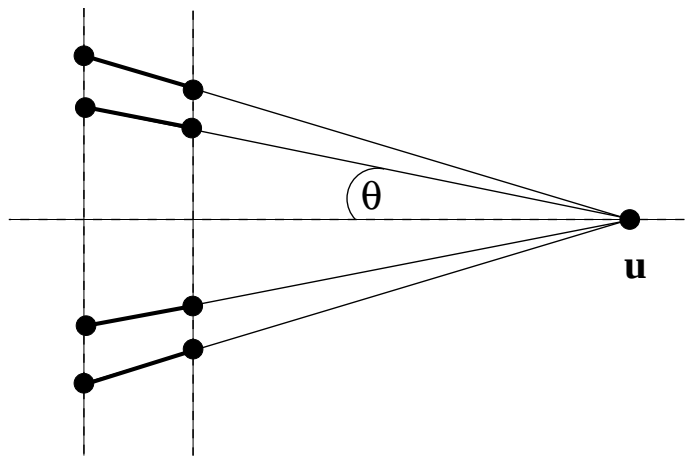
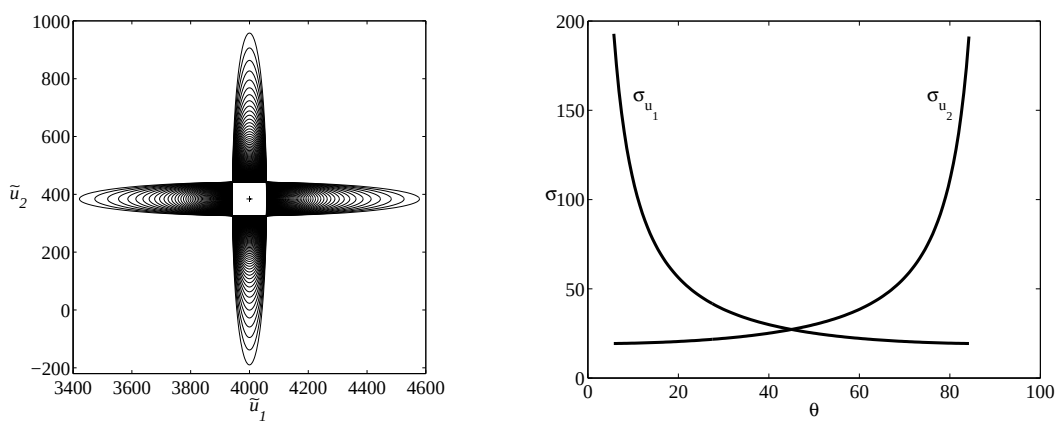


Figure 3.26: The configuration investigating the effect of line spacing on the magnitude and orientation of the covariance of the intersection point.



(a) (b)

Figure 3.27: Covariance in the intersection of two lines as a function of relative line orientation. (a) Covariance ellipses for a set of line pairs with a changing relative angle. The covariance ellipses with horizontal major axes are for lines close to parallel, with the major axis decreasing in size and then extending in the vertical direction as θ increases beyond 45°. (b) Standard deviations for the vertical ($\sigma_{\tilde{u}_1}$) and horizontal ($\sigma_{\tilde{u}_2}$) components of the intersection point as a function of θ.

points on the ideal line. The constraints can be written for the known relative lengths of pairs of line segments, known angles and pairs of equal unknown angles.

An unstratified approach to rectification has been developed using the conic dual to the circular points, a conic representation of both circular points and the vanishing line. The conic is linearly constrained by orthogonal line segments on the world plane. Particular attention has been devoted to the case of rectification of a rectangle of unknown aspect ratio, where three of the four degrees of freedom of the rectification homography are available, and the associated ambiguity. The rectangle case has important application to images of buildings, where rectangular structures on the facades are plentiful. It will be shown in the following chapter how the aspect ratio ambiguity can be resolved where there is sufficient information to compute internal parameters of the camera.

Finally, a Maximum Likelihood estimate of the point of intersection of a set of nominally coincident line segments under noisy measurement conditions has been developed. This is very important for plane rectification because of the reliance on vanishing points computed from the intersection of imaged parallel lines. It also proves important in the following chapters on single and two view calibration, where vanishing points are vital and image features are obtained from the intersection of boundary edges.

A first order error analysis of line intersection has been performed and found to be a good estimate under the noise assumptions. Synthetic experiments on the covariance of a point computed from line intersection suggests that the uncertainty in the computed point behaves much as would be expected. Increasing the number of lines reduces the uncertainty, and the direction of the uncertainty depends on the relative orientation of the line segments.

Chapter 4

Calibration and Reconstruction in One View

> It is, I think, agreed by all, that *distance* of itself, and immediately, cannot be seen. For *distance* being a line directed end-wise to the eye, it projects only one point in the fund of the eye. Which point remains invariably the same, whether the distance be longer or shorter.
>
> Berkeley, *An Essay Towards a New Theory of Vision, 1731*

4.1 Introduction

This chapter details methods of using the ubiquitous carpentered world geometry to calibrate cameras and compute 3D information from a single view. There are several reasons why camera calibration is important in this context.

1. Known internal calibration allows complete rectification of planes given only their vanishing line.

2. With known internal calibration and the vanishing line of a plane, the orientation of the plane relative to the camera may be computed.

3. With internal calibration and *two* vanishing lines the relative orientation between two planes can be computed.

Thus internal calibration together with the plane rectification techniques of the previous chapter allow 3D reconstruction of objects such as buildings from a single view *without world measurements.*

Information from scene structure and partial knowledge of the camera parameters are combined to perform the calibration. The chapter proceeds by describing three sources of single view calibration constraints:

1. the vanishing points of orthogonal directions,

2. metric rectified planes,

3. knowledge of some of the internal camera parameters.

All constraints are expressed in terms of the IAC. This approach has the considerable advantage that the constraints are *linear* equations in the elements of the IAC. The constraints from different sources may thus be combined effortlessly. Furthermore, it is shown that the often used constraints of known aspect ratio and (zero) skew are geometrically equivalent to the constraints arising from a metric plane. This in turn leads to constructions for the combination of constraints. These constructions prove particularly useful in the degeneracy analysis that follows.

Section 4.2 describes the applicable constraints. Section 4.3 examines the frequently encountered carpentered world case: the vanishing points of three orthogonal directions are visible in an image taken with a camera with unit aspect ratio and zero skew. Degenerate conditions are addressed in detail. Section 4.4 goes on to describe a first order error analysis and the use of additional constraints in the three vanishing point scenario.

A novel camera calibration object is described in 4.5. It is created by placing squares (or other planar objects with known geometry) on two or more planes in the image. The rectification of these planes calibrates the camera without any knowledge of 3D measurements, and the planes do not need to be orthogonal. In section 4.6 it is shown how the rectangle ambiguity described in

section 3.4 is resolved with calibration. The specific calibration scenarios that appear in the chapter are summarized in table 4.1.

Section 4.7 draws the calibration and rectification techniques together and describes reconstruction from a single view with several examples.

	Scene constraints	Internal parameter constraints
1.	Vanishing points of three orthogonal directions	Square pixels
2.	Vanishing points of three orthogonal directions	Square pixels and known principal point
3.	Three metric scene planes	None

Table 4.1: Single view calibration configurations investigated in this chapter.

4.2 Constraints on the IAC

This section describes single view calibration constraints as geometric constructions involving the IAC and points and lines in the image. The linearity of the constraints becomes evident when the expressions are expanded in the unknown elements of ω.

The algebraic representation of the IAC is a symmetric matrix with elements

$$\omega = \begin{pmatrix} \omega_1 & \omega_2 & \omega_4 \\ \omega_2 & \omega_3 & \omega_5 \\ \omega_4 & \omega_5 & \omega_6 \end{pmatrix} \tag{4.1}$$

Linear expressions in the elements of ω are written in the form

$$\boldsymbol{\kappa}^{\mathsf{T}} \boldsymbol{\omega}_{\mathsf{v}} = 0$$

where $\boldsymbol{\omega}_{\mathsf{v}} = (\omega_1, \ldots, \omega_6)^{\mathsf{T}}$ is the vector of the elements of ω and $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_6)^{\mathsf{T}}$ is the vector of coefficients.

The sections following will show how these constraints may be combined in various situations.

4.2.1 Vanishing points of orthogonal directions

The vanishing points of directions orthogonal in space have a special relationship to the IAC – the points are conjugate with respect to ω (see figure 4.1). For vanishing points $\mathbf{u}$ and $\mathbf{v}$, conjugacy is expressed

$$\mathbf{u}^\mathsf{T}\omega\mathbf{v} = 0 \tag{4.2}$$

This is a thinly disguised statement of the fact that the back-projected rays of the vanishing points have a dot product of zero, and may be derived as follows. The vanishing points $\mathbf{u}$ and $\mathbf{v}$ are the projections of points at infinity in 3D. Let these points be

$$\mathbf{U} = (U_1,\ U_2,\ U_3,\ 0)^\mathsf{T} = (\tilde{\mathbf{U}}^\mathsf{T},\ 0)^\mathsf{T} \text{ and } \mathbf{V} = (V_1,\ V_2,\ V_3,\ 0)^\mathsf{T} = (\tilde{\mathbf{V}}^\mathsf{T},\ 0)^\mathsf{T}$$

and, since they are orthogonal,

$$\tilde{\mathbf{U}}^\mathsf{T}\tilde{\mathbf{V}} = U_1V_1 + U_2V_2 + U_3V_3 = 0$$

Points $\mathbf{U}$ and $\mathbf{V}$ are projected to image points $\mathbf{u}$ and $\mathbf{v}$ by projection matrix $\mathsf{P} = \mathsf{K}(\mathsf{R}|\mathbf{t})$ to

$$\begin{aligned}\mathbf{u} &= \mathsf{P}\mathbf{U} = \mathsf{K}\mathsf{R}\tilde{\mathbf{U}} \text{ and}\\ \mathbf{v} &= \mathsf{P}\mathbf{V} = \mathsf{K}\mathsf{R}\tilde{\mathbf{V}}\end{aligned}$$

Now

$$\begin{aligned}\tilde{\mathbf{U}}^\mathsf{T}\tilde{\mathbf{V}} &= (\mathsf{R}^\mathsf{T}\mathsf{K}^{-1}\mathbf{u})^\mathsf{T}(\mathsf{R}^\mathsf{T}\mathsf{K}^{-1}\mathbf{v})\\ &= \mathbf{u}^\mathsf{T}\mathsf{K}^{-\mathsf{T}}\mathsf{R}\mathsf{R}^\mathsf{T}\mathsf{K}^{-1}\mathbf{v}\\ &= \mathbf{u}^\mathsf{T}\mathsf{K}^{-\mathsf{T}}\mathsf{K}^{-1}\mathbf{v} = \mathbf{u}^\mathsf{T}\omega\mathbf{v}\end{aligned}$$

The linearity of the constraint on ω is clear on expanding (4.2). Writing $\mathbf{u} = (u_1,\ u_2,\ u_3)^\mathsf{T}$ and

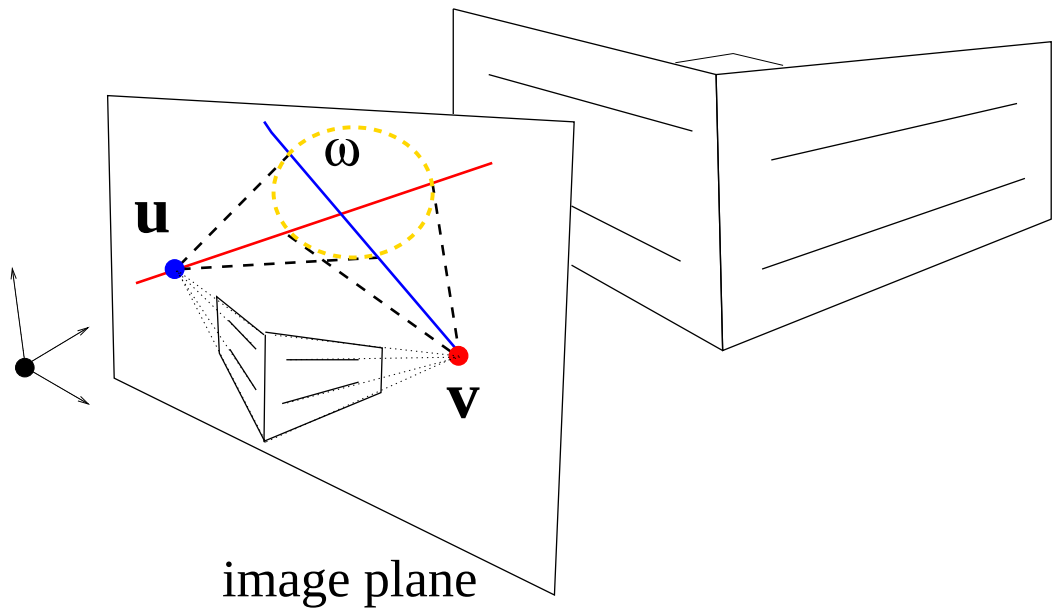


Figure 4.1: The vanishing points of orthogonal directions are conjugate with respect to ω. For example, the parallel lines on two orthogonal scene planes are imaged with vanishing points $\mathbf{u}$ and $\mathbf{v}$. Then $\mathbf{u}$ lies on the polar of $\mathbf{v}$ with respect to ω and $\mathbf{v}$ on the polar of $\mathbf{u}$.

$\mathbf{v} = (v_1,\ v_2,\ v_3)^\mathsf{T}$, (4.2) takes the form

$$\begin{aligned} &u_1v_1\omega_1 + (u_1v_2 + u_2v_1)\omega_2 + u_2v_2\omega_3 \\ &+(u_1v_3 + u_3v_1)\omega_4 + (u_2v_3 + u_3v_2)\omega_5 + u_3v_3\omega_6 = 0 \end{aligned} \tag{4.3}$$

In vector form, with the coefficient vector

$$\boldsymbol{\kappa}_{\mathbf{uv}} = (u_1v_1,\ u_1v_2 + u_2v_1,\ u_2v_2,\ u_1v_3 + u_3v_1,\ u_2v_3 + u_3v_2,\ u_3v_3)^\mathsf{T}$$

(4.3) becomes

$$\boldsymbol{\kappa}_{\mathbf{uv}}^\mathsf{T}\boldsymbol{\omega}_\mathsf{v} = 0 \tag{4.4}$$

For each pair of orthogonal vanishing points, an additional constraint of this form is obtained. Five such constraints determines ω and thence K. In the carpentered world, however, lines in *three*

orthogonal directions are commonly found. This case is examined in section 4.3.

4.2.2 Metric rectified planes

The imaged circular points of a plane lie on the IAC and thus satisfy

$$\mathbf{I}^{\mathsf{T}}\omega\mathbf{I} = 0 \text{ and } \mathbf{J}^{\mathsf{T}}\omega\mathbf{J} = 0 \tag{4.5}$$

As with the vanishing point property above, this relation may be simply derived. Consider a point on the absolute conic

$$\mathbf{X_I} = (X_{\mathbf{I}1},\ X_{\mathbf{I}2},\ X_{\mathbf{I}3},\ 0)^{\mathsf{T}} = (\tilde{\mathbf{X}}_{\mathbf{I}},\ 0)^{\mathsf{T}}$$

Since $\mathbf{X_I}$ lines on Ω_∞, it satisfies

$$\tilde{\mathbf{X}}_{\mathbf{I}}^{\mathsf{T}}\Omega_\infty\tilde{\mathbf{X}}_{\mathbf{I}} = \tilde{\mathbf{X}}_{\mathbf{I}}^{\mathsf{T}}\tilde{\mathbf{X}}_{\mathbf{I}} = 0$$

Point $\mathbf{X_I}$ is projected by $\mathsf{P} = \mathsf{K}(\mathsf{R}|\mathbf{t})$ to

$$\mathbf{I} = \mathsf{P}\mathbf{X_I} = \mathsf{K}\mathsf{R}\tilde{\mathbf{X}}_{\mathbf{I}}$$

So $\tilde{\mathbf{X}}_{\mathbf{I}} = \mathsf{R}^{\mathsf{T}}\mathsf{K}^{-1}\mathbf{I}$ and

$$\begin{aligned}\tilde{\mathbf{X}}_{\mathbf{I}}^{\mathsf{T}}\tilde{\mathbf{X}}_{\mathbf{I}} &= (\mathsf{R}^{\mathsf{T}}\mathsf{K}^{-1}\mathbf{I})^{\mathsf{T}}(\mathsf{R}^{\mathsf{T}}\mathsf{K}^{-1}\mathbf{I}) = \mathbf{I}^{\mathsf{T}}\mathsf{K}^{-\mathsf{T}}\mathsf{K}^{-1}\mathbf{I} \\ &= \mathbf{I}^{\mathsf{T}}\omega\mathbf{I}\end{aligned}$$

The circular points occur in complex conjugate pairs, and the above applies equally to $\mathbf{X_J} = \mathrm{conj}(\mathbf{X_I})$.

This property of the circular points also has a clear geometric interpretation mentioned in chapter 2. Any plane $\boldsymbol{\pi}$ intersects the plane at infinity $\boldsymbol{\pi}_\infty$ in a line. This line is the ideal line of $\boldsymbol{\pi}$, and it intersects the absolute conic in the circular points of $\boldsymbol{\pi}$, as illustrated in figure 4.2.

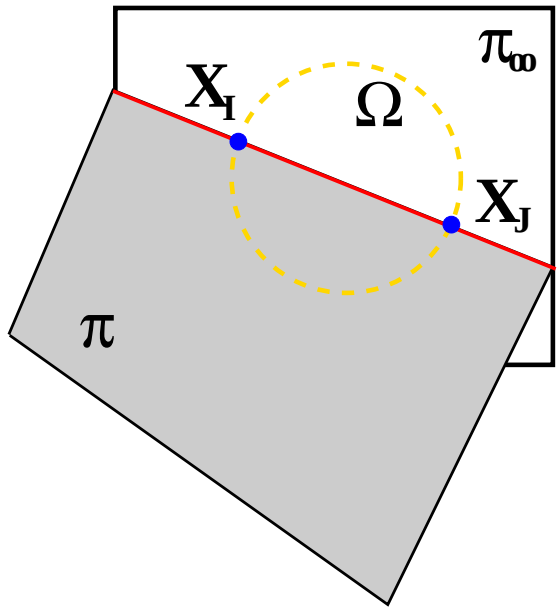


Figure 4.2: A plane $\boldsymbol{\pi}$ in space intersects the plane at infinity $\boldsymbol{\pi}_\infty$ in a line. This line is the line at infinity or ideal line of plane $\boldsymbol{\pi}$. The ideal lines intersects the absolute conic Ω_∞ in two points $\mathbf{X_I}$ and $\mathbf{X_J}$ – the circular points of the plane.

Ω_∞ projects to ω, the ideal line is imaged as the vanishing line $\mathbf{l}_\infty$ of the plane, and the circular points project to $\mathbf{I}$ and $\mathbf{J}$, as in figure 4.3. The vanishing line and imaged circular points define the metric rectification of the imaged plane (described in chapter 3), and thus if the rectification parameters for a scene plane are known, (4.5) provides two linear constraints on the elements of ω.

In practice, all the circular point information is contained in one of the complex conjugate points. Writing out the real and imaginary parts of *either* $\mathbf{I}^\top\omega\mathbf{I} = 0$ *or* $\mathbf{J}^\top\omega\mathbf{J} = 0$ yields two linear expressions in the elements of ω, denoted in vector form

$$\begin{aligned} \boldsymbol{\kappa}_R^\top \boldsymbol{\omega}_\mathsf{v} &= 0 \\ \text{and} \quad \boldsymbol{\kappa}_I^\top \boldsymbol{\omega}_\mathsf{v} &= 0 \end{aligned} \tag{4.6}$$

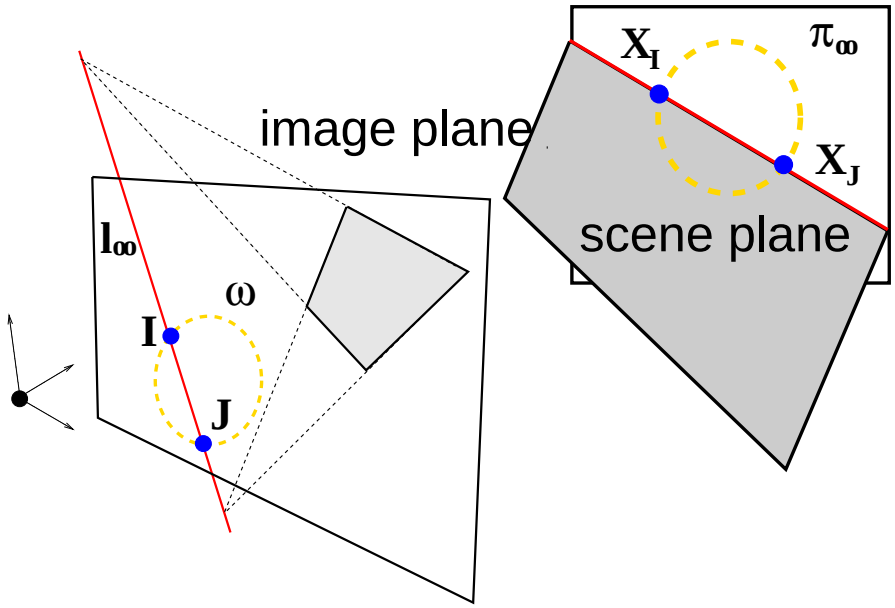


Figure 4.3: Projection of Ω_∞ to ω, $\mathbf{X_I}$ and $\mathbf{X_J}$ to $\mathbf{I}$ and $\mathbf{J}$ and the ideal line of $\boldsymbol{\pi}$ to $\mathbf{l}_\infty$. Knowing the rectification parameters for the image of $\boldsymbol{\pi}$ determines two linear constraints from (4.5)

4.2.3 Internal camera parameters

The general five parameter internal camera model (2.4) may be simplified in many cases by *a priori* knowledge of some of the parameters. A number of common constraints from known internal parameters in various combinations are presented below.

Zero Skew

That camera skew is zero is the most commonly applied internal parameter constraint. It specifies that the vertical and horizontal imaging axes are orthogonal. This is a reasonable assumption for most cameras, although there are cases where it does not apply, for example if a photographic negative is enlarged, and the paper is not parallel to the plane of the negative.

The assertion that $k = 0$ can be applied by specifying that the vanishing points of the vertical

and horizontal axes are orthogonal, and thus conjugate with respect to ω[117]

$$\begin{aligned} (1,\ 0,\ 0)\ \omega\ (0,\ 1,\ 0)^{\mathsf{T}} &= 0 \\ \text{or simply} \qquad \omega_2 &= 0 \end{aligned}$$

Aspect ratio

It is possible to write a *quadratic* constraint on elements of ω when the aspect ratio of a general camera is known. By inspection of the elements of ω in (2.13), a known aspect ratio r yields the following constraint[117]

$$\frac{\omega_3}{\omega_1} - \frac{\omega_2^2}{\omega_1^2} = \frac{1}{r^2}$$

The constraint becomes linear, however, when the camera has zero skew and $\omega_2 = 0$. In this case, the constraint can be posed geometrically as the orthogonality of any pair of directions which are orthogonal only after transformation by K^{-1} :

$$\begin{aligned} (1,\ r,\ 0)\ \omega\ (1,\ -r,\ 0) &= 0 \\ \omega_1 - r^2\omega_3 &= 0 \end{aligned}$$

In practice CDD cameras can often be considered to have zero skew and unit (or known) aspect ratio – to have square pixels. It will now be shown that a square pixel camera provides metric plane constraints on the IAC identical to those of any scene plane.

The square pixel camera

A square pixel camera has the form

$$\mathsf{K} = \begin{pmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{pmatrix} \tag{4.7}$$

Since K exhibits isotropic scaling and has no skew factor, it is a similarity transformation. Now, points on a world plane coinciding with the image plane are projected to image co-ordinates by simply transforming by K. The circular points of the image plane, with canonical co-ordinates $(1,\ \pm i,\ 0)^\mathsf{T}$ are thus imaged to the points

$$\mathbf{I} = \mathsf{K}(1,\ i,\ 0)^\mathsf{T} \text{ and } \mathbf{J} = \mathsf{K}(1,\ -i,\ 0)^\mathsf{T}$$

But, since K is a similarity, these points are invariant to transformation by K and

$$\mathbf{I} = \mathsf{K}(1,\ i,\ 0)^\mathsf{T} = (1,\ i,\ 0)^\mathsf{T} \text{ and } \mathbf{J} = \mathsf{K}(1,\ -i,\ 0)^\mathsf{T} = (1,\ -i,\ 0)^\mathsf{T}$$

The imaged circular points of a plane are thus known, and can be used to constrain the IAC just as with any scene plane for which the imaged circular points are known. Expanding (4.5) and considering real and imaginary parts one obtains the expected equations

$$\begin{aligned} \omega_1 - \omega_3 &= 0 \\ \text{and} \quad \omega_2 &= 0 \end{aligned} \tag{4.8}$$

The circular point interpretation might appear superfluous, since these relations are obvious from the form of ω. The geometric approach does, however, pay dividends in what follows, particularly in section 4.3. The most important point is that there is no intrinsic geometric distinction between the constraints obtained from the knowledge that the camera has square pixels and metric information from an imaged scene plane.

Note that the square pixel formulation applies equally well to any camera with zero skew and known aspect ratio (that is, with rectangular pixels). The image plane circular points are still known, although they do not have their canonical co-ordinates, and linearly constrain ω.

Known principal point

The principal point is the intersection of the optic axis and the image plane. It is thus commonly located at the centre of the image, provided the image is not cropped. A known principal point provides two constraints on K and ω since u_0 and v_0 are known. Geometrically, the centre of the IAC is translated from the origin to the principal point $\mathbf{p}$

$$\mathbf{p} = \mathsf{K}(0,\ 0,\ 1)^\mathsf{T} = (u_0,\ v_0,\ 1)^\mathsf{T}$$

The centre of a conic is defined geometrically as its pole with respect to to the ideal line, which gives a pair of linear equations on ω from the homogeneous polarity

$$\omega\mathbf{p} = (0,\ 0,\ 1)^\mathsf{T} \tag{4.9}$$

The resulting linear equations are

$$\begin{aligned} u_0\omega_1 + v_0\omega_2 + \omega_4 &= 0 \\ u_0\omega_2 + v_0\omega_3 + \omega_5 &= 0 \end{aligned} \tag{4.10}$$

4.3 Three vanishing points and a square pixel camera

This section describes the calibration of a square pixel camera from a single image of a cuboid object. Since the carpentered world is replete with approximately cuboid objects, this is a powerful and widely applicable approach. The solution appeared in the the photogrammetry literature in the work of Gracie[41] and was later reproduced by Caprile and Torre[11].

In Gracie's approach, the vanishing points of three orthogonal directions are computed in the image from the intersection of parallel lines or known length ratios (as in chapter 3). The back-projected rays of the three vanishing points are expressed in terms of the principal point and focal length. These rays are orthogonal and thus their dot products are zero, giving formulas for the three internal parameters. A construction is also given: the principal point lies at the orthocentre of the

triangle of vanishing points.

The method presented here is much the same, but is extended in three ways

1. Constraints from the vanishing points are expressed in terms of the IAC.
2. Degeneracy conditions, where the camera parameters are not fully constrained, are analysed.
3. Errors in the image data (the vanishing points) and assumptions about camera parameters are propagated to give a covariance on the computed internal parameters.

4.3.1 Five constraints on ω

Using (4.2), the conjugacy constraint on ω from a pair of orthogonal vanishing points may be applied to each pair of vanishing points

$$\begin{aligned} \mathbf{u}^{\mathsf{T}}\omega\mathbf{v} &= 0 \\ \mathbf{u}^{\mathsf{T}}\omega\mathbf{z} &= 0 \\ \mathbf{v}^{\mathsf{T}}\omega\mathbf{z} &= 0 \end{aligned}$$

for vanishing points $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{z}$. This provides three independent linear constraints on ω. Since the camera has square pixels, the circular points of the image plane constrain ω from

$$\begin{aligned} \mathbf{I}^{\mathsf{T}}\omega\mathbf{I} &= 0 \\ \mathbf{J}^{\mathsf{T}}\omega\mathbf{J} &= 0 \end{aligned}$$

giving two more constraints.

There are thus five linear constraints on the five degrees of freedom of ω, which may be written in matrix form as

$$\mathsf{A}\boldsymbol{\omega}_{\mathsf{v}} = \begin{pmatrix} \boldsymbol{\kappa}_{\mathbf{uv}} & \boldsymbol{\kappa}_{\mathbf{uz}} & \boldsymbol{\kappa}_{\mathbf{vz}} & \boldsymbol{\kappa}_{I} & \boldsymbol{\kappa}_{R} \end{pmatrix}^{\mathsf{T}} \boldsymbol{\omega}_{\mathsf{v}} = \mathbf{0} \tag{4.11}$$

where

$$\mathsf{A} = \begin{pmatrix} u_1v_1 & u_1z_1 & v_1z_1 & 0 & 1 \\ u_1v_2+u_2v_1 & u_1z_2+u_2z_1 & v_1z_2+v_2z_1 & 1 & 0 \\ u_2v_2 & u_2z_2 & v_2z_2 & 0 & -1 \\ u_1v_3+u_3v_1 & u_1z_3+u_3z_1 & v_1z_3+v_3z_1 & 0 & 0 \\ u_2v_3+u_3v_2 & u_2z_3+u_3z_2 & v_2z_3+v_3z_2 & 0 & 0 \\ u_3v_3 & u_3z_3 & v_3z_3 & 0 & 0 \end{pmatrix}^{\mathsf{T}}$$

The first three columns of A^{T} are the coefficients in vectors $\boldsymbol{\kappa}_{\mathbf{uv}}$, $\boldsymbol{\kappa}_{\mathbf{uz}}$ and $\boldsymbol{\kappa}_{\mathbf{vz}}$, obtained from the three pairs of vanishing points. The two remaining columns are the constraints from the imaged circular points.

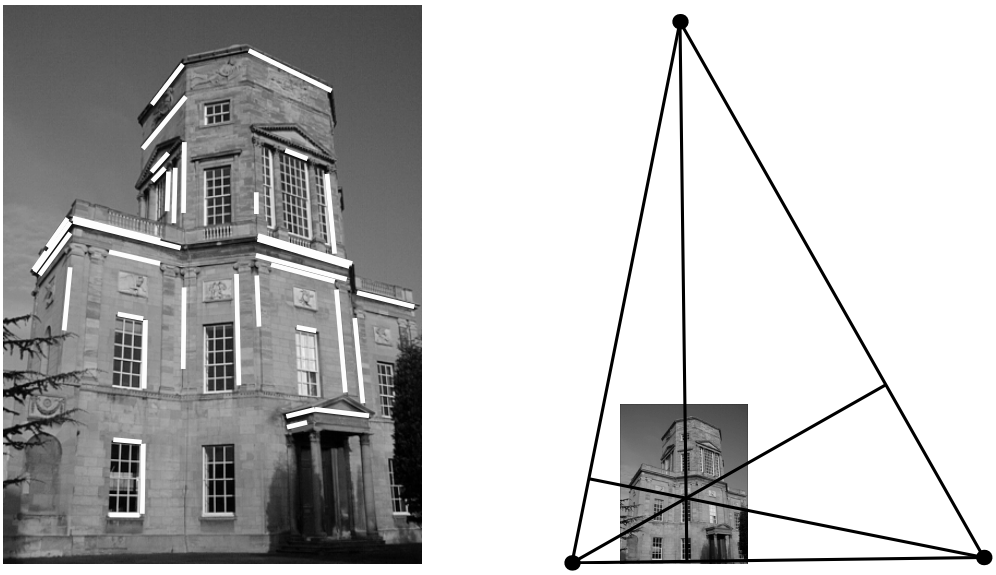
The solution $\boldsymbol{\omega}_{\mathsf{v}}$ is the null vector of A (if A is rank 5). The symmetric matrix ω can then be formed from $\boldsymbol{\omega}_{\mathsf{v}}$ and K computed by Cholesky decomposition. Figure 4.4 (a) shows an image of a building with lines in three orthogonal directions. The vanishing points of each of these three directions, shown in figure 4.4 (b), provide the three constraints on the internal parameters, and define the triangle with the principal point at its orthocentre.

It is now necessary to ask under what conditions these constraints are degenerate, where A will have rank less than five. The null space of A is then more than one dimensional, and the camera K is not uniquely defined. It is clear, for example, from the orthocentre construction, that if one of the vanishing points is ideal the orthocentre lies anywhere on the line joining the two finite vanishing points (see figure 4.10). The principal point is thus constrained to lie on a line. A general analysis of degeneracy follows.

4.3.2 Degeneracy analysis

It will be shown here that the internal calibration of a camera from the vanishing points of three orthogonal directions and metric knowledge of *any* world plane is degenerate *if the vanishing line of the world plane intersects any vanishing point.*

The general construction for the five constraints appears in figure 4.5. The three orthogonal



(a) (b)

Figure 4.4: Internal parameter estimation. (a) Image of the Radcliffe Observatory, Oxford, with sets of parallel line segments defining vanishing points of three orthogonal directions. (b) The triangle with the vanishing points as vertices. The principal point of the camera lies at the orthocentre of the triangle. f = 1048.6, u_0 = 398.8 and v_0 = 567.1. The image size is 768 by 1024.

vanishing points, which for notational convenience will be called $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$, are the vertices of a *self-polar triangle.* That is, since

$$\begin{aligned} \mathbf{v}_1^\mathsf{T}\omega\mathbf{v}_2 &= 0 \\ \mathbf{v}_1^\mathsf{T}\omega\mathbf{v}_3 &= 0 \\ \mathbf{v}_2^\mathsf{T}\omega\mathbf{v}_3 &= 0 \end{aligned} \tag{4.12}$$

each vanishing point is conjugate to the other two with respect to ω, and lies on the intersection of the polars of the other two with respect to ω. The polars $\mathbf{l}_1$, $\mathbf{l}_2$ and $\mathbf{l}_3$ are the sides of the triangle and satisfy

$$\mathbf{l}_i = \mathbf{v}_j \times \mathbf{v}_k = \omega\mathbf{v}_i$$

$$\text{and } \mathbf{v}_i^\mathsf{T}\mathbf{l}_j = \mathbf{v}_i^\mathsf{T}\mathbf{l}_k = 0$$

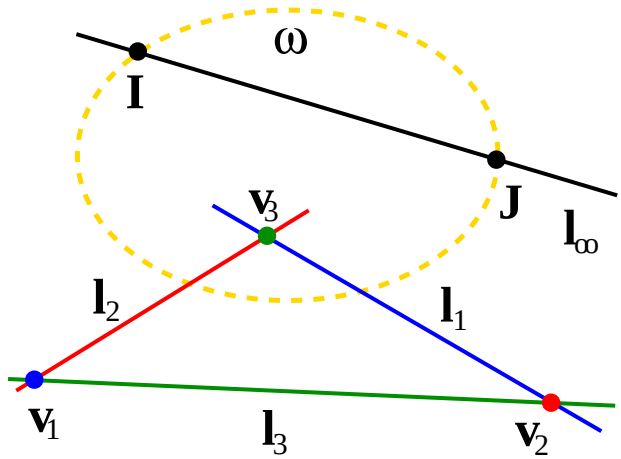


Figure 4.5: Construction for the constraints on ω obtained from the vanishing points of three orthogonal directions, $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$, and a metric plane with imaged circular points $\mathbf{I}$ and $\mathbf{J}$. Note that the construction is a representation of three mutual polarity relationships with respect to a *virtual* conic. It is thus not possible to draw the self-polar triangle showing all three polarities with tangent lines, since one of the poles ($\mathbf{v}_3$ in this case) appears inside the conic. Nevertheless, the diagram provides a useful graphical summary of the constraints.

The analysis proceeds by parameterizing the degrees of freedom remaining in ω once it is constrained by the vanishing points. The effect of adding a pair of imaged circular point constraints is then examined more closely.

Parameterizing ω

Given the three equations of (4.12), there are three constraints on the five degrees of freedom of ω. This leaves a two parameter family of conics that satisfy (4.12). Consider now the polar lines $\mathbf{l}_1$, $\mathbf{l}_2$ and $\mathbf{l}_3$ as degenerate (rank 1) point conics of the form $\mathbf{l}_i\mathbf{l}_i^\mathsf{T}$. For example, a point lying on the conic $\mathsf{L}_1 = \mathbf{l}_1\mathbf{l}_1^\mathsf{T}$ of this form satisfies $\mathbf{x}^\mathsf{T}\mathsf{L}_1\mathbf{x} = 0$. It is then possible to define a conic

$$\mathsf{C}(\mu,\nu,\eta) = \mu\mathbf{l}_1\mathbf{l}_1^\mathsf{T} + \nu\mathbf{l}_2\mathbf{l}_2^\mathsf{T} + \eta\mathbf{l}_3\mathbf{l}_3^\mathsf{T} \tag{4.13}$$

It is straight forward to verify that $\mathbf{v}_i^\mathsf{T}\mathsf{C}\mathbf{v}_j = 0\ \forall\ \mu,\ \nu,\ \eta$. Thus C satisfies the conjugacy constraints of the self-polar triangle with the two free parameters of the conic encoded by the homogeneous vector $(\mu,\nu,\eta)^\mathsf{T}$. We consider now how the imaged circular points constrain these two degrees of

freedom.

Imaged circular point constraints

Applying (4.5) to C

$$\begin{aligned} \mathbf{I}^\mathsf{T}\mathsf{C}\mathbf{I} &= 0 \\ \mathbf{J}^\mathsf{T}\mathsf{C}\mathbf{J} &= 0 \end{aligned} \tag{4.14}$$

In general, (4.14) determines ω by constraining the two degrees of freedom in $\mathsf{C}(\mu, \nu, \eta)$. The problem now is to find under what conditions the two parameter family of C is *not* fully constrained by (4.14). It will be shown that

Theorem 1 *Given a self-polar triangle with respect to ω and two imaged circular points on ω, ω is not fully determined if the line through the circular points on the conic intersects at least one vertex of the self-polar triangle.*

Proof The constraints in (4.14) are linear in the parameters $(\mu, \nu, \eta)^\mathsf{T}$ from (4.13), and may be written

$$\begin{bmatrix} (\mathbf{I}^\mathsf{T}\mathbf{l}_1)^2 & (\mathbf{I}^\mathsf{T}\mathbf{l}_2)^2 & (\mathbf{I}^\mathsf{T}\mathbf{l}_3)^2 \\ (\mathbf{J}^\mathsf{T}\mathbf{l}_1)^2 & (\mathbf{J}^\mathsf{T}\mathbf{l}_2)^2 & (\mathbf{J}^\mathsf{T}\mathbf{l}_3)^2 \end{bmatrix} \begin{pmatrix} \mu \\ \nu \\ \eta \end{pmatrix} = \mathsf{A}_\mathsf{c} \begin{pmatrix} \mu \\ \nu \\ \eta \end{pmatrix} = \mathbf{0} \tag{4.15}$$

In general this 2×3 matrix will have rank 2. However, if it drops rank the ambiguity in the conic is not resolved – there remains at least a one parameter family of conics satisfying all the constraints. Algebraically, the coefficient matrix A_c drops rank if all the 2×2 sub-matrices are singular. So, writing the determinants of all three sub-matrices of A_c:

$$d_{23} = (\mathbf{I}^\mathsf{T}\mathbf{l}_2)^2(\mathbf{J}^\mathsf{T}\mathbf{l}_3)^2 - (\mathbf{I}^\mathsf{T}\mathbf{l}_3)^2(\mathbf{J}^\mathsf{T}\mathbf{l}_2)^2 \tag{4.16}$$

$$d_{13} = (\mathbf{I}^\mathsf{T}\mathbf{l}_3)^2(\mathbf{J}^\mathsf{T}\mathbf{l}_1)^2 - (\mathbf{I}^\mathsf{T}\mathbf{l}_1)^2(\mathbf{J}^\mathsf{T}\mathbf{l}_3)^2 \tag{4.17}$$

$$d_{12} = (\mathbf{I}^\mathsf{T}\mathbf{l}_1)^2(\mathbf{J}^\mathsf{T}\mathbf{l}_2)^2 - (\mathbf{I}^\mathsf{T}\mathbf{l}_2)^2(\mathbf{J}^\mathsf{T}\mathbf{l}_1)^2 \tag{4.18}$$

The subscripts of d_{ij} are chosen to reflect which sides $\mathbf{l}_i$ and $\mathbf{l}_j$ of the self-polar triangle appear in the expressions for each determinant. The general form is

$$d_{ij} = (\mathbf{I}^\top \mathbf{l}_i)^2 (\mathbf{J}^\top \mathbf{l}_j)^2 - (\mathbf{I}^\top \mathbf{l}_j)^2 (\mathbf{J}^\top \mathbf{l}_i)^2$$

Now, by difference of squares

$$d_{ij} = \left[(\mathbf{I}^\top \mathbf{l}_i)(\mathbf{J}^\top \mathbf{l}_j) - (\mathbf{I}^\top \mathbf{l}_j)(\mathbf{J}^\top \mathbf{l}_i)\right] \left[(\mathbf{I}^\top \mathbf{l}_i)(\mathbf{J}^\top \mathbf{l}_j) + (\mathbf{I}^\top \mathbf{l}_j)(\mathbf{J}^\top \mathbf{l}_i)\right] \tag{4.19}$$

By the *identity of Lagrange* [63]

$$(\mathbf{a}^\top \mathbf{c})(\mathbf{b}^\top \mathbf{d}) - (\mathbf{a}^\top \mathbf{d})(\mathbf{b}^\top \mathbf{c}) = (\mathbf{a} \times \mathbf{b})^\top (\mathbf{c} \times \mathbf{d})$$

the first term in (4.19) becomes

$$(\mathbf{I}^\top \mathbf{l}_i)(\mathbf{J}^\top \mathbf{l}_j) - (\mathbf{I}^\top \mathbf{l}_j)(\mathbf{J}^\top \mathbf{l}_i) = (\mathbf{I} \times \mathbf{J})^\top (\mathbf{l}_i \times \mathbf{l}_j) = \mathbf{l}_\infty^\top \mathbf{v}_k$$

where $\mathbf{l}_\infty = \mathbf{I} \times \mathbf{J}$ is the vanishing line of the metric plane providing the circular point constraints. The second term can be written in the form of a degenerate rank 2 point conic consisting of the lines $\mathbf{l}_i$ and $\mathbf{l}_j$

$$(\mathbf{I}^\top \mathbf{l}_i)(\mathbf{J}^\top \mathbf{l}_j) + (\mathbf{I}^\top \mathbf{l}_j)(\mathbf{J}^\top \mathbf{l}_i) = \mathbf{I}^\top (\mathbf{l}_i \mathbf{l}_j^\top + \mathbf{l}_j \mathbf{l}_i^\top) \mathbf{J} = \mathbf{I}^\top \mathsf{L}_{ij} \mathbf{J}$$

Thus (4.19) may be rewritten as

$$d_{ij} = \left[\mathbf{l}_\infty^\top \mathbf{v}_k\right] \left[\mathbf{I}^\top \mathsf{L}_{ij} \mathbf{J}\right] = f_{ij} g_{ij} \tag{4.20}$$

The term f_{ij} is zero if the line $\mathbf{l}_\infty$ through $\mathbf{I}$ and $\mathbf{J}$ intersects the vertex $\mathbf{v}_k$ of the self-polar triangle. The term g_{ij} is zero if $\mathbf{I}$ and $\mathbf{J}$ are conjugate with respect to the conic L_{ij}. The conditions under which these occur will now be described. There are three cases to consider; $\mathbf{l}_\infty$ intersecting two vanishing points of the self-polar triangle, intersecting one vanishing point of the self-polar

triangle and no intersection. For the sake of clarity, the general subscripts will be replaced by numbers, referring directly to $\mathbf{v}_1$, $\mathbf{v}_2$ and so on. Since the problem is symmetric in these quantities, there is no loss of generality.

Case 1: l_∞ intersects two vanishing points.

When the line l_∞ intersects two vanishing points, say $\mathbf{v}_1$ and $\mathbf{v}_2$, $d_{13} = 0$ and $d_{23} = 0$ because $f_{13} = 0$ and $f_{23} = 0$. But, $f_{12} \neq 0$ since $\mathrm{l}_\infty = \mathrm{l}_3$ and $\mathbf{v}_3$ are not incident (see figure 4.6). Now it will be shown that $\mathbf{I}^\mathsf{T}\mathsf{L}_{12}\mathbf{J} = 0$ and thus $g_{12} = 0$. The equality $\mathbf{I}^\mathsf{T}\mathsf{L}_{12}\mathbf{J} = 0$ is true if $\mathbf{I}$ and $\mathbf{J}$ are conjugate with respect to the degenerate conic L_{12}. From section 2.2.2, they are conjugate if the two pairs of points $\mathbf{v}_1$ and $\mathbf{v}_2$ and $\mathbf{I}$ and $\mathbf{J}$ are harmonic. Now, since $\mathbf{v}_1$ and $\mathbf{v}_2$ are the vanishing points of orthogonal directions and $\mathbf{I}$ and $\mathbf{J}$ are the imaged circular points of their common plane, the two point pairs are harmonic and $\mathbf{I}^\mathsf{T}\mathsf{L}_{12}\mathbf{J} = 0$.

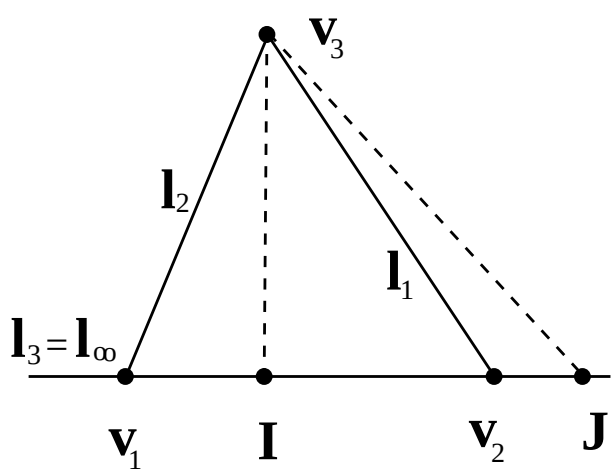


Figure 4.6: Construction for case 1, l_∞ intersects $\mathbf{v}_1$ and $\mathbf{v}_2$.

Case 2: l_∞ intersects one vanishing point.

When the line l_∞ intersects one vanishing point, say $\mathbf{v}_1$, $d_{23} = 0$ because $f_{23} = 0$. But, $f_{12} \neq 0$ and $f_{13} \neq 0$ since l_∞ does not intersect $\mathbf{v}_2$ or $\mathbf{v}_3$ (see figure 4.7). It will now be shown that $\mathbf{I}^\mathsf{T}\mathsf{L}_{12}\mathbf{J} = 0$ and thus $g_{12} = 0$. By symmetry, this also applies to $\mathbf{I}^\mathsf{T}\mathsf{L}_{13}\mathbf{J} = 0$. Again, the expression $\mathbf{I}^\mathsf{T}\mathsf{L}_{12}\mathbf{J} = 0$ is true if $\mathbf{I}$ and $\mathbf{J}$ are conjugate with respect to the degenerate conic L_{12}. The same considerations as in case 1 apply because any point $\mathbf{p}$ on l_1 is the vanishing point of a direction orthogonal to $\mathbf{v}_1$.

Therefore, if point $\mathbf{p}$ is the intersection of l_∞ and l_1, point pairs $\mathbf{I}$, $\mathbf{J}$, and $\mathbf{v}_1$, $\mathbf{p}$ are harmonic and $\mathbf{I}$ and $\mathbf{J}$ are conjugate with respect to L_{12}.

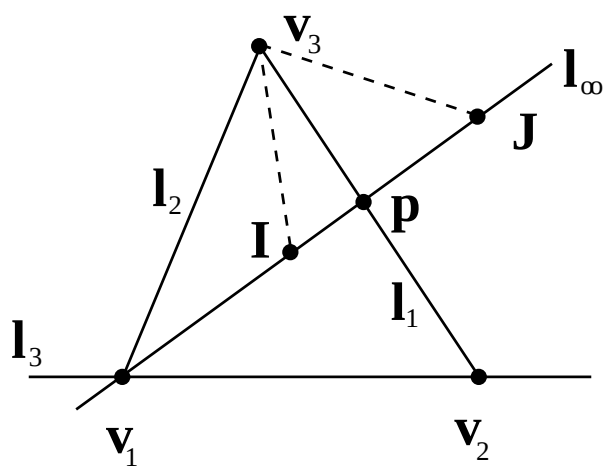


Figure 4.7: Construction for case 2, l_∞ intersects $\mathbf{v}_1$.

Case 3: l_∞ does not intersect any vanishing points.

When the line l_∞ does not intersect any vanishing points, none of the terms f_{ij} are zero. Also, l_∞ intersects the self-polar triangle in two points not at the vertices, $\mathbf{p}$ and $\mathbf{q}$ (see figure 4.8). These two points are thus not the vanishing points of orthogonal directions and the points $\mathbf{I}$, $\mathbf{J}$, $\mathbf{p}$ and $\mathbf{q}$ are not harmonic. None of the f_{ij} terms vanish and A_c is full rank. This completes the proof ■

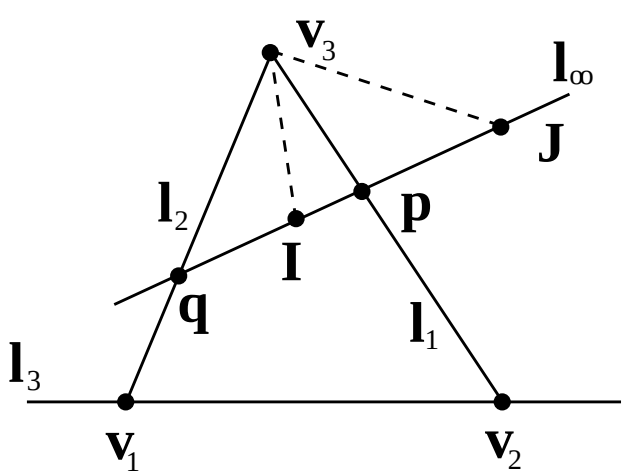


Figure 4.8: Construction for case 3, l_∞ does not intersect any vanishing points.

The one parameter family

It has been shown that, given three orthogonal vanishing points and the imaged circular points of one metric plane (usually from a square pixel camera), a degeneracy exists when the vanishing line of the metric plane passes through one of the vanishing points. Only 4 independent constraints on ω exist when this degeneracy occurs, leaving a 1 parameter family of conics satisfying the self-polar triangle conjugacy and circular point pair constraints. A parameterization of this family is now given.

Assume that the imaged circular points $\mathbf{I}$ and $\mathbf{J}$ lie on the line $\mathbf{l}_\infty$, which intersects the triangle of vanishing points $\mathbf{v}_1\mathbf{v}_2\mathbf{v}_3$ in the vertex $\mathbf{v}_1$. Referring to figure 4.9, construct the lines $\mathbf{l}_{\mathbf{I}2} = \mathbf{I} \times \mathbf{v}_2$, $\mathbf{l}_{\mathbf{J}2} = \mathbf{J} \times \mathbf{v}_2$, $\mathbf{l}_{\mathbf{I}3} = \mathbf{I} \times \mathbf{v}_3$ and $\mathbf{l}_{\mathbf{J}3} = \mathbf{J} \times \mathbf{v}_3$. These four lines can be considered as two degenerate (rank 2) conics $\mathsf{L}_2 = \mathbf{l}_{\mathbf{I}2}\mathbf{l}_{\mathbf{J}2}^\top + \mathbf{l}_{\mathbf{J}2}\mathbf{l}_{\mathbf{I}2}^\top$ and $\mathsf{L}_3 = \mathbf{l}_{\mathbf{I}3}\mathbf{l}_{\mathbf{J}3}^\top + \mathbf{l}_{\mathbf{J}3}\mathbf{l}_{\mathbf{I}3}^\top$. It will be shown below that the one parameter family of proper conics

$$\mathsf{G}(\tau, \rho) = \tau\mathsf{L}_1 + \rho\mathsf{L}_2$$

satisfies all 5 constraints under consideration i.e. the conjugacy constraints $\mathbf{v}_i^\top\mathsf{G}\mathbf{v}_j = 0$ and circular point pair constraints $\mathbf{I}^\top\mathsf{G}\mathbf{I} = \mathbf{J}^\top\mathsf{G}\mathbf{J} = 0\ \forall\tau, \rho$.

The circular point pair constraints are clearly satisfied by G since $\mathbf{I}$ and $\mathbf{J}$ lie on both conics L_1 and L_2. The conjugacy constraints must each be considered for L_2 and L_3. Firstly, $\mathbf{v}_1^\top\mathsf{L}_2\mathbf{v}_2 = 0$ and $\mathbf{v}_3^\top\mathsf{L}_2\mathbf{v}_2 = 0$, since $\mathbf{v}_2^\top\mathbf{l}_{\mathbf{I}2} = 0$ and $\mathbf{v}_2^\top\mathbf{l}_{\mathbf{J}2} = 0$. Similarly $\mathbf{v}_1^\top\mathsf{L}_3\mathbf{v}_3 = 0$ and $\mathbf{v}_2^\top\mathsf{L}_3\mathbf{v}_3 = 0$.

It remains only to show that $\mathbf{v}_1^\top\mathsf{L}_2\mathbf{v}_3 = 0$ and $\mathbf{v}_1^\top\mathsf{L}_3\mathbf{v}_2 = 0$. This follows from the same reasoning as in the proof of Theorem 1, since point pairs $\mathbf{I}$, $\mathbf{J}$,and $\mathbf{v}_1$, $\mathbf{v}_3$ are harmonic, $\mathbf{v}_1$ and $\mathbf{v}_3$ are conjugate with respect to the conic consisting of the lines $\mathbf{I} \times \mathbf{v}_2$ and $\mathbf{J} \times \mathbf{v}_2$. The same applies to $\mathbf{v}_1^\top\mathsf{L}_3\mathbf{v}_2 = 0$.

In the case that $\mathbf{l}_{\mathbf{IJ}}$ intersects *two* vertices of triangle, say $\mathbf{v}_1$ and $\mathbf{v}_2$, $\mathbf{v}_2 \times \mathbf{I} = \mathbf{v}_2 \times \mathbf{J} = \mathbf{l}_3$ and $\mathsf{L}_2 = \mathbf{l}_3\mathbf{l}_3^\top$ and the parameterization still holds.

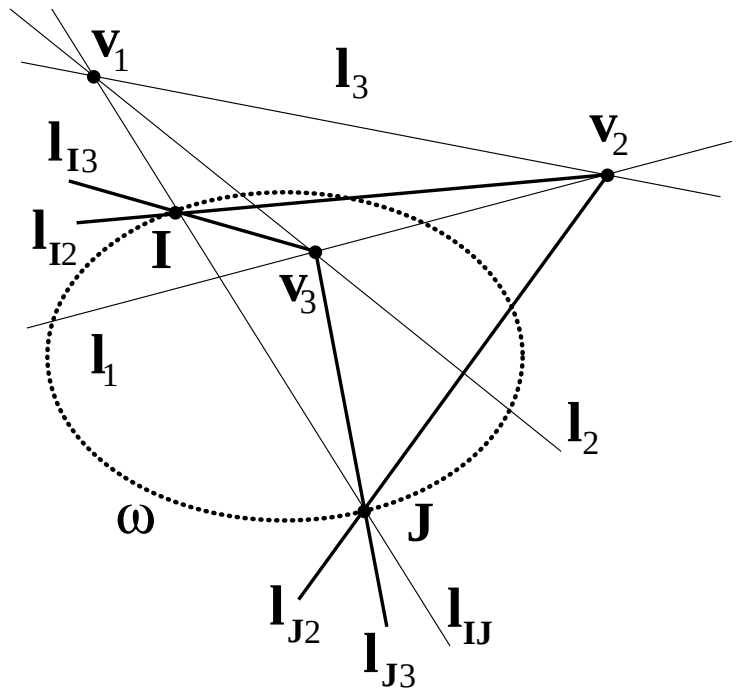


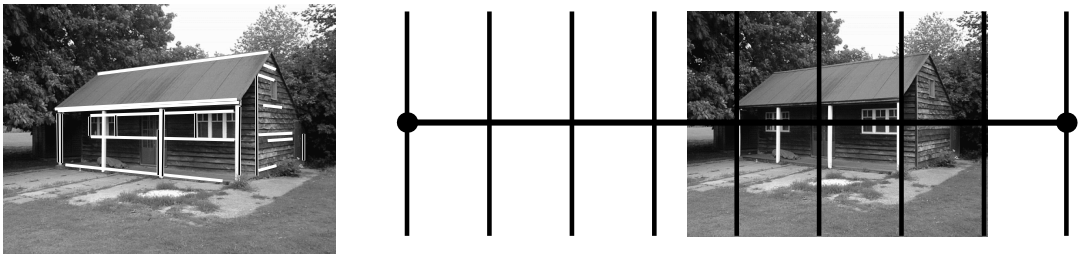
Figure 4.9: Construction for the one parameter family of conics

The square pixel camera degeneracy cases

The particular cases of the degeneracy arising with three vanishing points and a square pixel camera occur when one or two of the vanishing points are ideal. The image plane is the metric plane providing circular point constraints, and its vanishing line is the ideal line $(0,\ 0,\ 1)^{\mathsf{T}}$). Thus, when an imaged vanishing point is ideal, it lies on the vanishing line of the metric plane, and is collinear with the image plane circular points.

One ideal vanishing point: This is common in photographs of buildings when the camera is held so that the image plane is vertical. Vertical scene lines are parallel to the image plane and their vanishing point is imaged at infinity. An example appears in figure 4.10. The orthocentre of the triangle of vanishing points, the principal point, is defined anywhere on the line through the two finite vanishing points – see figure 4.10 (b). The coefficient matrix A from (4.11) is rank four.

Two ideal vanishing points: Two vanishing points imaged at infinity is less common, but can occur when the image plane is parallel to a scene plane, as in figure 4.11. There is again



(a) (b)

Figure 4.10: A single view of a building with the vanishing point of the vertical direction imaged at infinity. (a) Imaged parallel line segments in three orthogonal directions. (b) The finite vanishing points and lines in the pencil through the orthocentre. The principal point lies on the horizontal line shown through the two finite vanishing points.

a one parameter family of cameras, parametrized by unknown focal length. This follows from the observation that if two of three orthogonal vanishing points are ideal, the third must be the vanishing point of the optic axis of the camera, which is the principal point. Again, the coefficient matrix A is rank four.

Figure 4.11: Image of a doorway where the vanishing points of the vertical and horizontal directions are ideal. The finite vanishing point is at the principal point of the camera.

A graphical explanation

The ambiguity in internal camera parameters computed from a cuboid object and a metric plane may also be understood by considering the dual relationship between camera and structure. The degrees

of freedom remaining in the incompletely calibrated camera are associated with transformations that may be applied to the structure. This will become clearer with the example that follows.

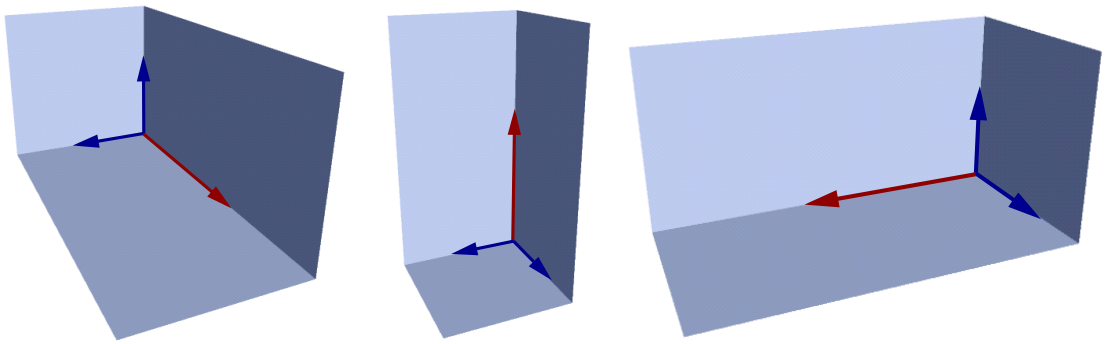
Consider figure 4.12, which shows an image of a cuboid building and an abstraction of the cuboid structure. The orthogonality of the three directions of the cuboid axes provides the three constraints on the camera from the self-polar triangle of vanishing points. The two degrees of freedom remaining in the camera parameters are equivalent to the transformations that can be applied to the structure while preserving orthogonality, as in figure 4.13. Specifically, ignoring similarity transformations, the structure can be scaled by different amounts in the three orthogonal directions. This transformation has two degrees of freedom (ignoring global scale).

(a) (b)

Figure 4.12: (a) Fellows quad, Merton College, Oxford. The dominant planes of the building are represented by three orthogonal planes in (b).

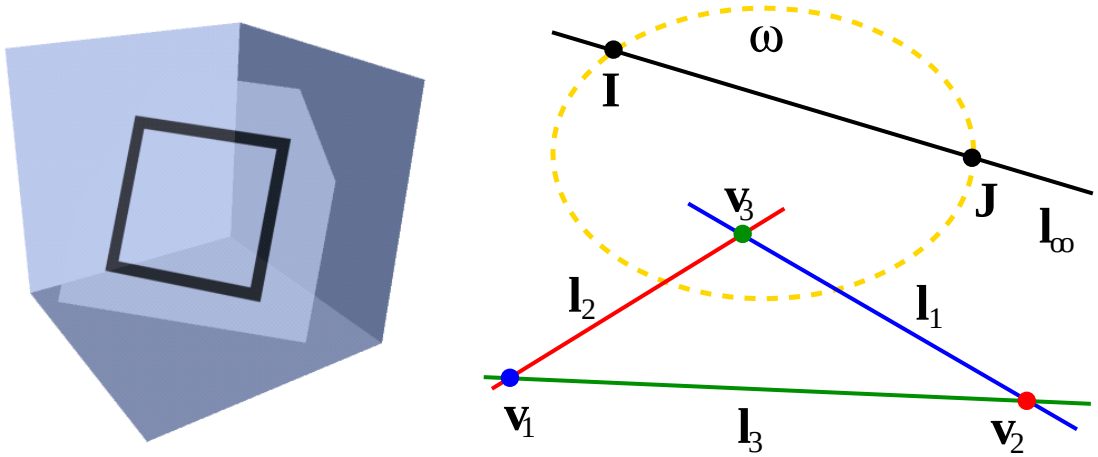
In the general case, the vanishing line of the metric plane does not intersect any vertices of the self-polar triangle. Structurally, a metric scene plane is defined, indicated in figure 4.14 (a) by a plane with a square on it. There is no transformation that will preserve the orthogonality of the cuboid plane as well as the structure of the metric plane. The construction on ω appears in figure 4.14 (b).

The degenerate case where the vanishing line of the metric plane intersects two vertices of the self-polar triangle appears in figure 4.15. In this case the metric scene plane is parallel to one of the cuboid planes. *Arbitrary* scaling of the structure in the direction orthogonal to the metric plane will preserve the metric plane structure *and* the orthogonality of the cuboid planes. This is the



(a) (b) (c)

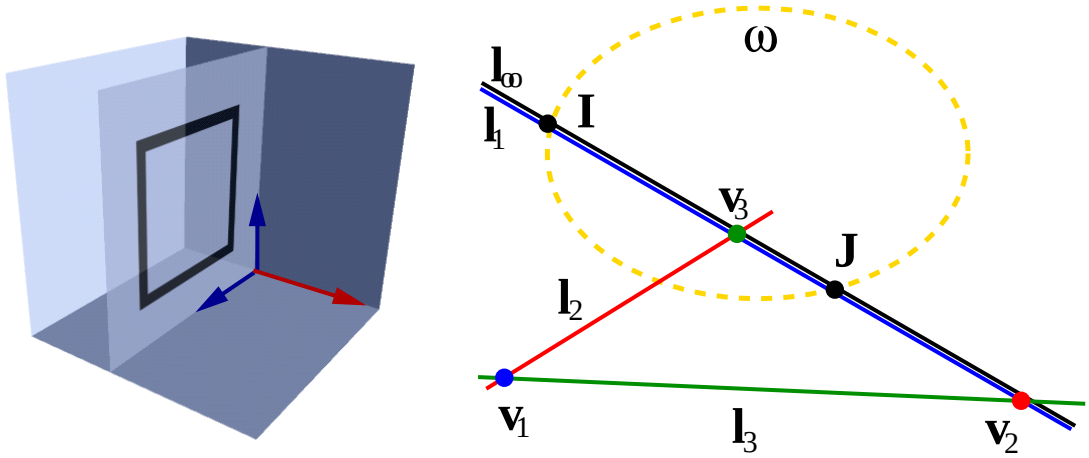
Figure 4.13: Examples of transformations that preserve orthogonality. There are two degrees of freedom in transformations of the structure that preserve orthogonality.



(a) (b)

Figure 4.14: (a) A general metric plane constraint. No combination of scaling as in figure 4.13 can preserve the orthogonality of the cuboid planes and the metric plane structure. (b) Construction on ω.

remaining degree of freedom. Figure 4.15 (b) shows the construction, with the vanishing line of the metric plane coinciding with one side of the self-polar triangle.



(a) (b)

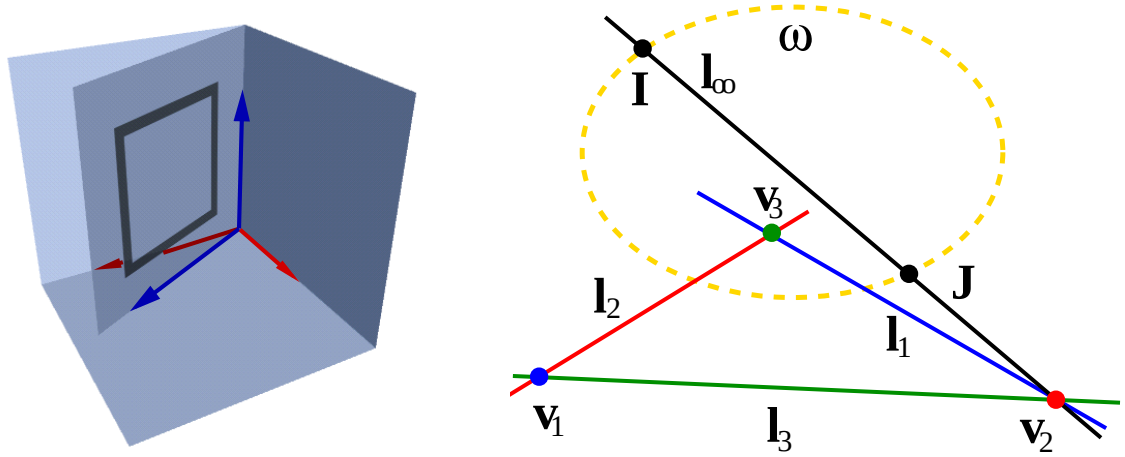
Figure 4.15: (a) The degeneracy occurring when the metric plane is parallel to one of the cuboid planes. (b) The construction with the vanishing line of the metric plane coinciding with a side of the self-polar triangle.

Finally, the second degenerate case occurs when the metric scene plane has one direction in common with the cuboid axes, as in figure 4.16. Any scaling *orthogonal* to the common direction with resultant equal to the scaling *in* the common direction preserves cuboid orthogonality and metric plane structure – the one degree of freedom in the calibration. The construction of figure 4.16 (b) shows the imaged condition, the vanishing line of the metric plane intersecting one vertex of the self-polar triangle.

The degenerate cases are summarized in table 4.2.

Incidence condition for any metric plane	Image plane is metric plane	Independent constraints on ω
$\mathbf{l}_\infty$ intersects no vanishing points	all vanishing points finite	5
$\mathbf{l}_\infty$ intersects one vanishing point	one vanishing point ideal	4
$\mathbf{l}_\infty$ intersects two vanishing points	two vanishing points ideal	4

Table 4.2: The number of independent constraints on ω for three vanishing points and the circular points of a metric plane, for both a general scene plane and the image plane.



(a) (b)

Figure 4.16: (a) The degeneracy occurring when the metric plane has one direction in common with the cuboid. (b) The construction with the vanishing line of the metric plane intersecting a vertex of the self-polar triangle.

4.4 Uncertainty and additional constraints

The uncertainty in camera parameters computed from three vanishing points and square pixels is quantified in this section. Using a first order error propagation the covariance matrix for camera parameters is developed and extended to the addition of a known principal point constraint. Before addressing the computations, a number of comments are in order.

There is a distinction to be made between hard and soft constraints on the internal parameters of a camera. For example, specifying that a camera has zero skew by writing $\omega_2 = 0$ is a soft constraint on ω. Explicitly parametrizing ω with elements ω_2 as zero is a hard constraint. A hard constraint ensures that the computed camera has a zero entry in the second column of the first row, while the soft constraint of skew zero on ω can, in the over-constrained case, result in a camera with a (usually small) non-zero skew.

When it comes to combining constraints from camera parameters with scene constraints, the choice of hard or soft constraints effects the computation required. For a square pixel camera, applying the zero skew and unit aspect ratio conditions as hard constraints yields the parametrization of (4.7). The three constraints from the vanishing points of orthogonal directions can be expanded for this form of K to give closed form expressions for f, u_0 and v_0. In general, however, there are

several advantages to soft constraints:

1. Soft constraints may be combined with other constraints and weighted by relative confidence.

2. Uncertainty in the specified internal parameters can be included in error analysis.

3. Degeneracy analysis is simplified by the consistent representation of different constraint types, as in the previous section.

In terms of the error analysis of the square pixel camera case, using soft constraints makes computing a Jacobian more difficult, although the result is the same. For hard constraints, direct differentiation of the closed form expressions yields a Jacobian for the three internal parameters, and thus covariance. Here, however, the full form of K is retained, and the skew and aspect ratio constrained using the image plane circular points. The solution for ω is then the null vector solution of (4.11). Since $\boldsymbol{\omega}_{\mathsf{v}}$ is defined by an implicit function, the Jacobian cannot be computed directly and requires the method outlined in the following section. The extra effort, however, is justified for two reasons. First, the method is trivially extended to the over-constrained case, such as when a known principal point with an associated prior uncertainty is introduced. Second, the over-constrained estimate of K has non-zero skew and non-unit aspect ratio, and covariances on these parameters can be computed.

Introducing a known principal point constraint to a square pixel camera in the presence of the vanishing points of three orthogonal directions is often useful. The degenerate condition that frequently occurs in images of buildings where the vertical vanishing point is close to ideal means that the principal point is poorly constrained. The principal point of most digital cameras, however, lies close to the centre of the image. How close is dependent on the manufacturing process, but in many cases the constraint that the principal point is at the centre may be applied as long as the image is not cropped with unknown parameters. Applying the constraint is simple. The two equations (4.10) are added to (4.11) by adding two rows to the coefficient matrix. The algebraic minimum given by the singular vector associated with the smallest singular value of the coefficient matrix provides a solution for $\boldsymbol{\omega}_{\mathsf{v}}$. The Jacobian follows directly from the five constraint case, with the addition of the derivatives with respect to the two additional variable (u_0 and v_0). An example of the degenerate

case with and without the principal point constraints, and including the error analysis, appears in section 4.4.2.

4.4.1 Error propagation

A first order error analysis of the square pixel configuration is presented in this section. The five constraint approach is analyzed assuming Gaussian error on the endpoints of the imaged parallel line segments defining vanishing points. The resulting covariance matrix describes the uncertainty in the computed camera parameters. Monte Carlo simulation is used to verify that the first order approximation is valid at reasonable levels of noise, and the analysis is shown to be consistent with the degeneracy results of the previous section.

The error analysis introduces two new components

1. Covariance of the null vector solution of (4.11), the vector of elements of ω.

2. Covariance of the camera parameters, the elements of K, given the elements of ω and the associated covariance matrix. This requires a first order error propagation through the Cholesky decomposition of a three by three matrix.

The null vector

The covariance of the vector $\boldsymbol{\omega}_{\mathsf{v}}$ obtained as the null vector in (4.11) may be computed using a general approach described by Faugeras[31]. The description below follows the treatment by Clarke[14]. The method provides an estimate of the Jacobian of the singular vector associated with the smallest singular value of the coefficient matrix A in (4.11). Using the SVD based approach allows easy extension to over constrained cases.

Given the constraint equation

$$\mathsf{A}\boldsymbol{\omega}_{\mathsf{v}} = \mathbf{0}$$

the algebraic minimum is found by considering the problem to be a constrained minimization of the

cost function subject to the constraint that the homogeneous vector $\boldsymbol{\omega}_{\mathsf{v}}$ has norm 1:

$$\min_{\boldsymbol{\omega}_{\mathsf{v}}} \mathcal{C} = \|\mathsf{A}\boldsymbol{\omega}_{\mathsf{v}}\| \quad \text{subject to} \qquad \|\boldsymbol{\omega}_{\mathsf{v}}\| = 1$$

The solution is given by the vector corresponding to the smallest singular value of A. This may equivalently be expressed as the eigenvector corresponding to the smallest eigenvalue of the matrix $\mathsf{A}^{\mathsf{T}}\mathsf{A}$

$$\boldsymbol{\omega}_{\mathsf{v}}^{*} = \mathsf{A}^{\mathsf{T}}\mathsf{A}\boldsymbol{\omega}_{\mathsf{v}} = \mu\boldsymbol{\omega}_{\mathsf{v}}$$

where $\boldsymbol{\omega}_{\mathsf{v}}^{*}$ is the eigenvector and μ the smallest eigenvalue. With five constraints on the calibration, the solution is exact and $\mu = 0$. The problem now is to find the Jacobian $\frac{\partial \boldsymbol{\omega}_{\mathsf{v}}}{\partial \mathbf{d}}$, where

$$\mathbf{d} = (\mathbf{u}^{\mathsf{T}},\ \mathbf{v}^{\mathsf{T}},\ \mathbf{z}^{\mathsf{T}})^{\mathsf{T}}$$

for the three vanishing points of orthogonal directions $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{z}$ each computed from the images of a set of parallel lines.

The difficulty is that $\boldsymbol{\omega}_{\mathsf{v}}$ is defined by an implicit function, so it is not possible to simply compute the derivative expressions. Faugeras appeals to the implicit function theorem from analysis to show that for the function

$$\Phi(\mathbf{d}, \boldsymbol{\omega}_{\mathsf{v}}) = \Phi(\mathbf{d}, \mathbf{f}(\mathbf{d})) = 0$$

a function $\mathbf{f}$ is defined for which the Jacobian

$$\mathsf{J}_{\boldsymbol{\omega}_{\mathsf{v}}} = \frac{\partial \boldsymbol{\omega}_{\mathsf{v}}}{\partial \mathbf{d}} = \frac{\partial \mathbf{f}(\mathbf{d})}{\partial \mathbf{d}} = -\left(\frac{\partial \Phi}{\partial \boldsymbol{\omega}_{\mathsf{v}}}\bigg|_{\mathbf{d}=\mathbf{d}_0}\right)^{-1} \frac{\partial \Phi}{\partial \mathbf{d}}\bigg|_{\boldsymbol{\omega}_{\mathsf{v}}=\boldsymbol{\omega}_{\mathsf{v}0}}$$

can be computed. Hence, writing

$$\Phi = \mathsf{A}^{\mathsf{T}}\mathsf{A}\boldsymbol{\omega}_{\mathsf{v}} \qquad \text{and} \qquad \frac{\partial \Phi}{\partial \boldsymbol{\omega}_{\mathsf{v}}} = \mathsf{A}^{\mathsf{T}}\mathsf{A}$$

the first term in the Jacobian expression is approximated by the pseudo-inverse

$$(\frac{\partial \Phi}{\partial \boldsymbol{\omega}_{\mathsf{v}}})^{-1} \approx (\mathsf{A}^{\mathsf{T}}\mathsf{A})^{+}$$

The second term is computed directly as follows. Each element of $\mathsf{A}^{\mathsf{T}}\mathsf{A}\boldsymbol{\omega}_{\mathsf{v}}$ can be written $\boldsymbol{\omega}_{\mathsf{v}}^{\mathsf{T}}\mathbf{a}_i$ where $\mathbf{a}_i = (a_{i1}, \ \ldots \ , \ a_{i6})^{\mathsf{T}}$ is the ith row of $\mathsf{A}^{\mathsf{T}}\mathsf{A}$. So,

$$\frac{\partial \Phi_i}{\partial \mathbf{d}} = \boldsymbol{\omega}_{\mathsf{v}}^{\mathsf{T}} \frac{\partial \mathbf{a}_i}{\partial \mathbf{d}}$$

and, defining matrices

$$\mathsf{B} = \begin{pmatrix} \boldsymbol{\omega}_{\mathsf{v}}^{\mathsf{T}} & & \\ & \boldsymbol{\omega}_{\mathsf{v}}^{\mathsf{T}} & \\ & & \ddots \end{pmatrix} \quad \text{and} \quad \mathsf{C} = \begin{pmatrix} \frac{\partial \mathbf{a}_1}{\partial \mathbf{d}} \\ \frac{\partial \mathbf{a}_2}{\partial \mathbf{d}} \\ \vdots \end{pmatrix}$$

$$\frac{\partial \Phi_i}{\partial \mathbf{d}} = \mathsf{BC}$$

The Jacobian is then

$$\mathsf{J}_{\boldsymbol{\omega}_{\mathsf{v}}} = -(\frac{\partial \Phi}{\partial \boldsymbol{\omega}_{\mathsf{v}}})^{-1} \frac{\partial \Phi}{\partial \mathbf{d}} = -(\mathsf{A}^{\mathsf{T}}\mathsf{A})^{+}\mathsf{BC}$$

and the covariance of $\boldsymbol{\omega}_{\mathsf{v}}$, $\Lambda_{\boldsymbol{\omega}_{\mathsf{v}}}$, is given by

$$\Lambda_{\boldsymbol{\omega}_{\mathsf{v}}} = (\mathsf{A}^{\mathsf{T}}\mathsf{A})^{+}\mathsf{BC}\Lambda_{\mathbf{d}}\mathsf{C}^{\mathsf{T}}\mathsf{B}^{\mathsf{T}}(\mathsf{A}^{\mathsf{T}}\mathsf{A})^{+} \tag{4.21}$$

where

$$\Lambda_{\mathbf{d}} = \begin{pmatrix} \Lambda_{\mathbf{u}} & & \\ & \Lambda_{\mathbf{v}} & \\ & & \Lambda_{\mathbf{z}} \end{pmatrix}$$

Cholesky decomposition

The Cholesky decomposition of the n by n symmetric positive definite matrix Z determines a lower triangular matrix L such that $\mathsf{Z} = \mathsf{L}\mathsf{L}^{\mathsf{T}}$. The elements of L are usually defined recursively[63]. Writing the element of row i and column j of L as L_{ij}, and similarly for the elements of Z

$$\begin{aligned}
L_{11} &= \sqrt{Z_{11}} \\
L_{jj} &= \sqrt{Z_{jj} - \sum_{s=1}^{j-1} L_{js}^2} && j = 2, \cdots, n \\
L_{j1} &= \frac{Z_{j1}}{L_{11}} && j = 2, \cdots, n \\
L_{jk} &= \frac{1}{L_{kk}}(Z_{jk} - \sum_{s=1}^{k-1} L_{js}L_{ks}) && j = k+1, \cdots, n; k \geq 2
\end{aligned} \tag{4.22}$$

Since $\omega = \mathsf{K}^{-\mathsf{T}}\mathsf{K}^{-1}$, substituting $\mathsf{K}^{-\mathsf{T}} = \mathsf{L}$ and $\omega = \mathsf{Z}$ gives

$$\mathsf{K}^{-1} = \begin{pmatrix} \sqrt{\omega_1} & \frac{\omega_2}{\sqrt{\omega_1}} & \frac{\omega_4}{\sqrt{\omega_1}} \\ 0 & \sqrt{\omega_3 - \frac{\omega_2^2}{\omega_1}} & \frac{\omega_5 - \frac{\omega_2\omega_4}{\omega_1}}{\sqrt{\omega_3 - \frac{\omega_2^2}{\omega_1}}} \\ 0 & 0 & \sqrt{\omega_6 - \frac{\omega_4^2}{\omega_1} - \frac{(\omega_5 - \frac{\omega_2\omega_4}{\omega_1})^2}{\omega_3 - \frac{\omega_2^2}{\omega_1}}} \end{pmatrix} \tag{4.23}$$

Inverting and normalizing by K_{33}, the internal calibration elements can be written as a vector

$$\begin{aligned}
\mathbf{k} &= \frac{1}{K_{33}}(K_{11},\ K_{12},\ K_{22},\ K_{13},\ K_{23})^{\mathsf{T}} \\
&= (f,\ k,\ rf,\ u_0,\ v_0)^{\mathsf{T}} \\
&= \frac{1}{\omega_1\omega_3 - \omega_2^2} \begin{pmatrix} \frac{1}{\sqrt{\omega_1}}\sqrt{(\omega_1\omega_3 - \omega_2^2)(\omega 1\omega_3\omega_6 - \omega_2^2\omega_6 - \omega_3\omega_4^2 - \omega_1\omega_5^2 + 2\omega_2\omega_4\omega_5)} \\ \frac{-\omega_2}{\sqrt{\omega_1}}\sqrt{\omega 1\omega_3\omega_6 - \omega_2^2\omega_6 - \omega_3\omega_4^2 - \omega_1\omega_5^2 + 2\omega_2\omega_4\omega_5} \\ \sqrt{\omega_1(\omega 1\omega_3\omega_6 - \omega_2^2\omega_6 - \omega_3\omega_4^2 - \omega_1\omega_5^2 + 2\omega_2\omega_4\omega_5)} \\ \omega_2\omega_5 - \omega_3\omega_4 \\ \omega_2\omega_4 - \omega_1\omega_5 \end{pmatrix}
\end{aligned}$$

The Jacobian $\mathsf{J}_{\mathbf{k}}^{\mathsf{T}} = \frac{\partial \mathbf{k}}{\partial \boldsymbol{\omega}_{\mathsf{v}}}$ then follows from straight forward differentiation. The elements of $\mathsf{J}_{\mathbf{k}}^{\mathsf{T}}$ are written out in full in appendix A

Now, given the covariance matrix of the elements of ω, $\Lambda_{\boldsymbol{\omega}_{\mathsf{v}}}$, the covariance of the internal camera parameters is

$$\Lambda_{\mathbf{k}} = \mathsf{J}_{\mathbf{k}} \Lambda_{\boldsymbol{\omega}_{\mathsf{v}}} \mathsf{J}_{\mathbf{k}}^{\mathsf{T}} \tag{4.24}$$

Uncertainty example and Monte Carlo simulation

It is necessary to make some assumptions about the noise in the image measurements. For the sake of demonstration, an isotropic Gaussian uncertainty with a standard deviation of 0.5 pixels is assumed on the endpoints of the imaged parallel line segments. The internal parameter covariance $\Lambda_{\mathbf{k}}$ is computed by first determining the covariance of $\boldsymbol{\omega}_{\mathsf{v}}$ using (4.21) and then applying (4.24).

Returning to the image used as an example in section 4.3.1, figure 4.4, the computed camera parameters and $\Lambda_{\mathbf{k}}$ appear in table 4.3 below. Since the camera is not over-constrained, the skew and aspect ratio are perfectly recovered, and there is no uncertainty in their values. Also, three standard deviations of the focal length is approximately 40 pixels. This is a reasonable uncertainty given the assumed image noise.

	Camera parameters	Covariance matrix				
f	1048.6	176.8862	0	176.8862	136.0478	6.7529
k	0	0	0	0	0	0
rf	1048.6	176.8862	0	176.8862	136.0478	6.7529
u_0	398.8	136.0478	0	136.0478	308.1398	-36.3323
v_0	567.1	6.7529	0	6.7529	-36.3323	94.7298

Table 4.3: The camera parameters and first order covariance matrix for the example appearing in figure 4.4, using the measured image date and assuming a 0.5 pixel standard deviation on the measured endpoints of imaged parallel line segments.

The validity of this result is verified by computing a Monte Carlo simulation from ideal data. The synthetic data is created by computing, as in section 3.6.2, a set of parallel line segments that do intersect in a unique point for each vanishing point. The Monte Carlo simulation proceeds by

adding random isotropic Gaussian noise to the endpoints of the synthetic line segments. The three vanishing points are computed and (4.11) is solved. The result of 10 000 iterations with line segment endpoint noise standard deviation of 0.5 pixels appears in table 4.4. Comparing this with table 4.5, the covariance from (4.21) and (4.24) applied to the synthetic data without noise, it is clear that there is very close agreement between the linear error propagation results and those computed from Monte Carlo simulation. The similarity is readily apparent in plots of the covariance of the principal point. Figure 4.17(a) shows the three standard deviation ellipse of the principal point computed from the synthetic data, and (b) the computed principal points and three standard deviation ellipse obtained from Monte Carlo simulation. The ellipses are plotted together in figure 4.17(c)

	Mean camera parameters	Covariance matrix				
f	1048.5	177.5390	0	177.5390	134.4944	6.5635
k	0	0	0	0	0	0
rf	1.0485	177.5390	0	177.5390	134.4944	6.5635
u_0	398.7	134.4944	0	134.4944	308.5633	-41.4693
v_0	567.4	6.5635	0	6.5635	-41.4693	96.5144

Table 4.4: Monte Carlo simulation camera parameters and covariance matrix for synthetic data extracted from the example in figure 4.4. Random Gaussian noise with a standard deviation of 0.5 pixels is added to the synthetic data in 10 000 iterations.

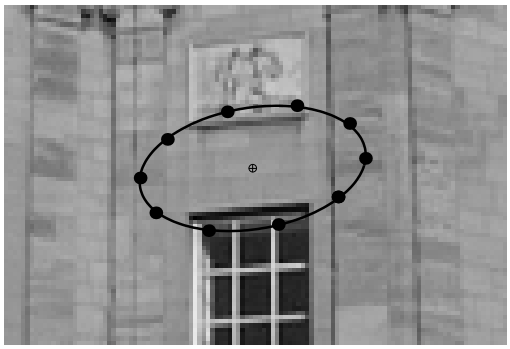
	Camera parameters	Covariance matrix				
f	1048.6	176.9172	0	176.9172	136.0746	6.7510
k	0	0	0	0	0	0
rf	1048.6	176.9172	0	176.9172	136.0746	6.7510
u_0	398.7	136.0746	0	136.0746	308.1869	-36.3393
v_0	567.1	6.7510	0	6.7510	-36.3393	94.7401

Table 4.5: The camera parameters and first order covariance matrix computed for synthetic data extracted from the example in figure 4.4.

4.4.2 Adding the constraint of a known principal point

The IAC is over constrained when the two constraints obtained from a known principal point are introduced. The result is a coefficient matrix with seven rows, and generally of rank five in the

(a) (b)



(c)

Figure 4.17: The three standard deviation ellipses of the principal point computed (a) from (4.21) and (4.24). (b) from Monte Carlo simulation. (b) also shows the principal points for all 10 000 iterations. (c) shows the two ellipses from the synthetic data (solid line) and Monte Carlo simulation (filled circles) simultaneously.

absence of noise. With noise, however, the rank increases to six and $\boldsymbol{\omega}_\mathsf{v}$ can be estimated by the singular vector associated with the smallest singular value. Note that, in terms of the associated Jacobian, the smallest eigenvalue of the constraint matrix is not zero, so the Jacobian is an approximation. Also, the relative scale of constraint rows in the coefficient matrix has an effect on the solution, so rows can in principle be weighted according to confidence. The examples which follow neglect such weighting, with reasonable results.

This section proceeds by giving an example of a near degenerate case of calibration from three vanishing points and a square pixel camera. The uncertainty in the camera parameters is shown to be consistent with the degeneracy analysis, and is then compared to the results obtained after adding the principal point constraints.

An example of a near ideal vertical vanishing point appears in figure 4.10. Table 4.6 shows the camera parameters and covariance matrix computed from (4.21) and (4.24) with data variances as in the last section. The three standard deviation ellipse of the principal point appears in figure 4.18. The variance for focal length and the horizontal component of the principal point are extremely large. The principal point is constrained to lie on the line connecting the finite vanishing points, in this case defining the horizontal line, and this is in fact the orientation of the major axis of the error ellipse.

Table 4.7 shows the camera parameters and covariance matrix computed from (4.11), (4.21) and (4.24), with (4.11) and (4.24) updated to take into account the principal point at the centre of the image. Standard deviations are as before; 0.5 pixels on endpoints of line segments, but with 10 pixels on u_0 and v_0. The three standard deviation ellipse of the principal point appears in figure 4.18. Monte Carlo simulation confirms the accuracy of these results. The tables are omitted for the sake of brevity.

Constraining the principal point to the centre of the image ensures that there are sufficient constraints on ω to estimate the camera in the presence of degenerate conditions. Note that the covariance of the principal point obtained from the error analysis is different from the prior covariance assumed on its co-ordinates. This reflects the combination of the prior uncertainty and the uncertainty associated with the other calibration constraints.

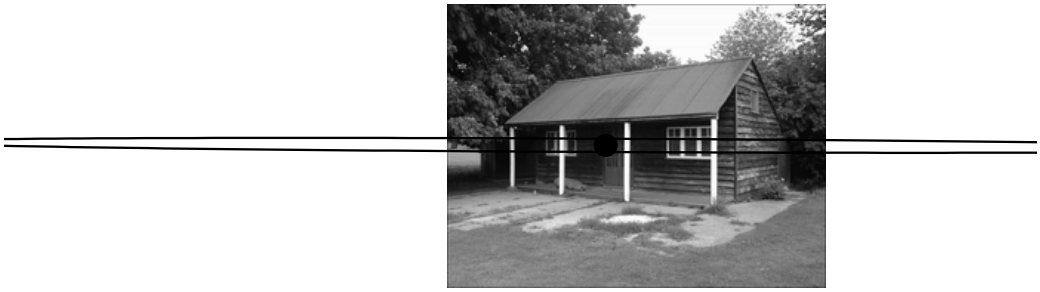


Figure 4.18: Principal point three standard deviation ellipses for the internal parameters computed from (4.11). The direction and extent of the ellipse agrees with the degeneracy analysis - the principal point is constrained to lie on the line joining the two finite vanishing points.

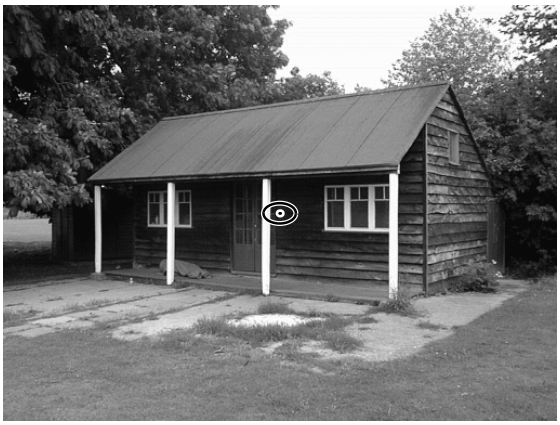


Figure 4.19: Principal point three standard deviation ellipse with the principal point constrained to lie at the centre of the image.

	Camera parameters	Covariance matrix				
f	1094.3	22662	0	22662	-94479	-453.5
k	0	0	0	0	0	0
rf	1094.3	22662	0	22662	-94479	-453.5
u_0	429.3	-94479	0	-94479	3.953e+05	1907.8
v_0	385.1	-453.5	0	-453.5	1907.8	45.548

Table 4.6: The camera parameters and first order covariance matrix for the case of one near ideal vanishing point.

	Camera parameters	Covariance matrix				
f	1071.3	97.4838	-0.1725	97.4833	-32.0640	0.3081
k	-0.0001	-0.1725	0.8822	-0.1743	-0.0840	0.2564
rf	1071.3	97.4833	-0.1743	98.6306	-32.0639	0.3253
u_0	512.0	-32.0640	-0.0840	-32.0639	99.9984	-0.0978
v_0	384.4	0.3081	0.2564	0.3253	-0.0978	39.2935

Table 4.7: The camera parameters and first order covariance matrix with one ideal vanishing point and the principal point constrained to lie at the centre of the image.

4.5 A three square calibration object

It was shown in section 4.2.2 that the imaged circular points of any world plane lie on ω and provide two constraints on ω. Chapter 3 went into some detail concerning the relationship between knowledge of the geometry of an imaged plane and the imaged circular points. The key point in this section is that knowing the metric geometry of the imaged plane is equivalent to knowing the imaged circular point parameters. It is thus clear that if the metric rectification of an imaged plane can be computed the imaged circular points are known, and thus two constraints on the camera internal parameters are available from (4.5).

The calibration object presented here exploits the metric scene plane constraints by capturing an image of three squares. Since the squares are simple to rectify, three pairs of imaged circular points are easily computed, giving six linear constraints on ω. An example and construction appear in figure 4.20, showing the intersection of ω with the vanishing lines of each plane in the imaged circular points. The main advantages of such a calibration object are that the imaged planes need not be orthogonal and the 3D co-ordinates of the calibration pattern need not be known. The only

constraint on the planes is that they are not coplanar, and so define three unique pairs of image circular points.

This method is closely related to the planar auto-calibration method of Triggs[106], but differs in that the imaged circular points of each plane are computed explicitly, leading to linear constraints on ω. A similar approach to constraining camera parameters from metric planes has been developed by both Sturm and Maybank[97], and Zhang[114]. Both present the constraints on ω obtained from a metric plane in terms of the columns of the world to image plane homography $\tilde{\mathsf{H}}$, writing

$$\begin{aligned} \mathbf{h}_1^\intercal \omega \mathbf{h}_2 &= 0 \\ \mathbf{h}_1^\intercal \omega \mathbf{h}_1 - \mathbf{h}_2^\intercal \omega \mathbf{h}_2 &= 0 \end{aligned} \tag{4.25}$$

for the first two columns $\mathbf{h}_1$ and $\mathbf{h}_2$ of $\tilde{\mathsf{H}}$. (Note that that $\tilde{\mathsf{H}}$, the homography used in these papers, is the inverse of the *rectifying homography* in chapter 3.) That these constraints are equivalent to imaged circular point constraints can be seen as follows.

The imaged circular points of a plane can be transferred from the world plane by $\tilde{\mathsf{H}}$ to give

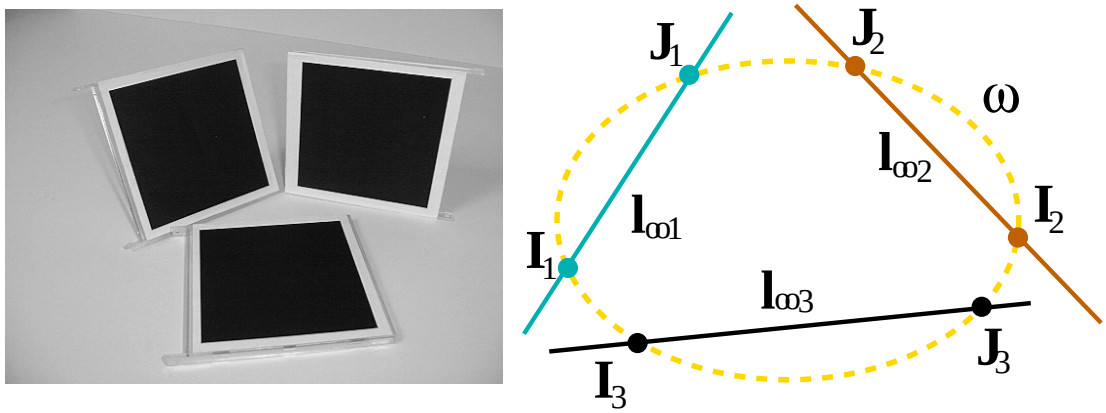
$$\mathbf{I} = \tilde{\mathsf{H}}(1,\ i,\ 0) = \mathbf{h}_1 + i\mathbf{h}_2$$

That $\mathbf{I}$ lies on ω is then

$$(\mathbf{h}_1 + i\mathbf{h}_2)^\intercal \omega (\mathbf{h}_1 + i\mathbf{h}_2) \quad = \quad \mathbf{h}_1^\intercal \omega \mathbf{h}_1 - \mathbf{h}_2^\intercal \omega \mathbf{h}_2 - 2i\mathbf{h}_1^\intercal \omega \mathbf{h}_2 = 0$$

Equating real and imaginary parts to zero yields (4.25).

Note that the three square configuration is chosen mainly for ease of use – a square pattern is easily created and rectified. Any plane for which metric properties can be computed can be used, such as patterns with rectangles or circles, or arbitrary patterns with known plane geometry. It is also possible to use fewer squares in combination with other constraints, such as the square pixel constraint and/or a known principal point. Patterns with multiple squares may also be used. The implementation here uses the orthogonality of the sides of the squares to compute rectification pa-



(a) (b)

Figure 4.20: A three square calibration object. (a) Image of three squares. (b) The six points on ω obtained from the imaged circular points of three world planes.

rameters (described in section 3.5). Since the edges of the imaged squares are used, long, unbroken edge are advantageous.

With three pairs of points giving six constraints, the calibration is over parametrized. A MLE solution is thus presented which minimizes the geometric error. A first order error propagation is computed and verified using Monte Carlo simulation.

4.5.1 Linear computation of ω

The circular points for each imaged square are determined directly from the orthogonality of the sides of the square and the diagonals. These constraints are sufficient to fully define the conic D_∞ dual to the circular points for each imaged plane using (3.17). The circular point parameters are then easily extracted from D by referring to (3.12). The resulting three imaged circular point pairs $\mathbf{I}_1,\ \mathbf{J}_1, \mathbf{I}_2,\ \mathbf{J}_2, \mathbf{I}_3,\ \mathbf{J}_3$, define six coefficient vectors $\boldsymbol{\kappa}_{R1},\ \boldsymbol{\kappa}_{I1},\ \boldsymbol{\kappa}_{R2},\ \boldsymbol{\kappa}_{I2},\ \boldsymbol{\kappa}_{R3},\ \boldsymbol{\kappa}_{I3}$ for six linear constraints on ω from (4.6)

$$\mathsf{A}\boldsymbol{\omega}_{\mathsf{v}} = (\boldsymbol{\kappa}_{R1},\ \boldsymbol{\kappa}_{I1},\ \boldsymbol{\kappa}_{R2},\ \boldsymbol{\kappa}_{I2},\ \boldsymbol{\kappa}_{R3},\ \boldsymbol{\kappa}_{I3})^{\mathsf{T}}\boldsymbol{\omega}_{\mathsf{v}} = \mathbf{0} \tag{4.26}$$

Since ω has five degrees of freedom, it is over-constrained. Under ideal conditions, A has rank five and the solution for ω is its null vector. Noise on the measurements, however, results in a

rank six coefficient matrix. The singular vector associated with the smallest singular value of A is the estimate of $\boldsymbol{\omega}_\mathsf{v}$ that minimizes $\|\mathsf{A}\boldsymbol{\omega}_\mathsf{v}\|$ subject to $\|\boldsymbol{\omega}_\mathsf{v}\| = 1$. As before, K follows from ω by Cholesky decomposition.

However, the ω computed by algebraic minimization will not exactly satisfy all the constraints. Geometrically, the imaged circular points constraining ω will not necessarily lie *on* the conic (see figure 4.21). The algebraic minimum can be improved by a Maximum Likelihood Estimate of ω (and the imaged circular points on ω) that minimizes the error in the measurements, the corners of the imaged squares. This is the subject of the next section.

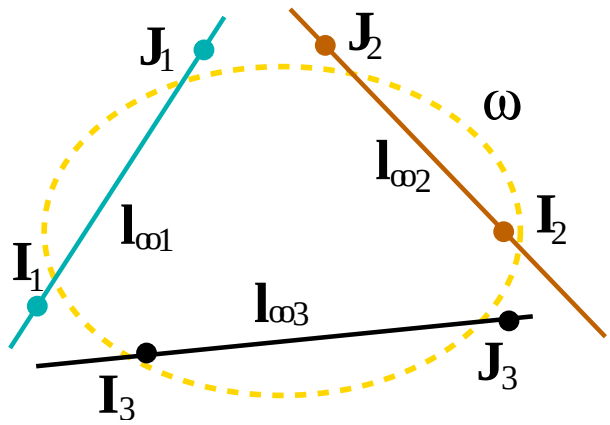


Figure 4.21: The six imaged circular points do not necessarily lie on the conic obtained from singular value decomposition of A.

4.5.2 A Maximum Likelihood Estimate of K

The goal of this section is to compute three pairs of imaged circular points that meet two criteria; they lie on the same conic and minimize reprojection error. The reprojection error is the Mahalanobis distance between the corners of the squares measured in the image and the corners computed by mappings of each of the world plane corners to the image. It is in the latter computation that the difficulty lies, since the mappings must be consistent with ω. The imaged circular points do not lie on the same conic, but each of the pairs encode a rectification homography that transforms an imaged square to a perfect world plane square. If these points are constrained to be consistent

with the computed ω, the rectification parameters defined by the new circular points will not rectify the imaged square to a square. This is illustrated in figure 4.22.

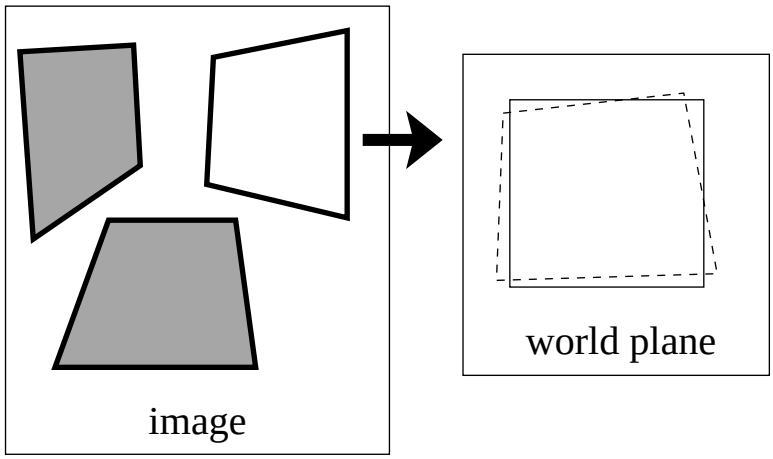


Figure 4.22: The imaged circular points of a particular imaged square encode a rectification homography that maps it to a perfect square (solid lines on the right). Any other pair of imaged circular points, constrained to lie on ω for example, define a homography that does *not* map the imaged square to a perfect square (dashed lines on the right).

The decomposition of the general projective transformation between a world plane and its image described in chapter 3 discards the similarity transformation to retain only the projective and affine components of the homography. To compute a reprojection error a transformation from world plane to image is required that, given a pair of imaged circular points, maps a world plane square *to the closest quadrilateral to the imaged square*. Applying a similarity transformation to the perfect square representing the world plane points before the transformation defined by the circular points makes this possible, since the transformation applied is then a general homography. The problem is to determine the similarity transformation as well as the rectification parameters for each square.

Including a similarity transformation increases the number of degrees of freedom associated with each square from four to eight.This leads to an iterative minimization scheme to compute the MLE of K using the Levenberg-Marquardt algorithm as follows:

1. Compute the linear estimate of K from (4.26) to initialize the minimization. This includes computing the rectification homographies $\mathsf{N}_i,\ i = 1, 2, 3,$ for each square, the transformations

that map the imaged squares to perfect squares on the world plane.

2. Construct the vector of co-ordinates of the 12 corners of the squares

$$\mathbf{e} = (x_1,\ y_1,\ \ldots\ ,\ x_{12},\ y_{12})^\mathsf{T}$$

The objective now is to compute a set of points $\hat{\mathbf{e}}$ such that the imaged squares defined by $\hat{\mathbf{e}}$ are rectified to perfect world squares by the three pairs of imaged circular points $\hat{\mathbf{I}}_i$ and $\hat{\mathbf{J}}_i$ for $i = 1, 2, 3$. Furthermore, $\hat{\mathbf{I}}_i$ and $\hat{\mathbf{J}}_i$ must all lie on the conic $\hat{\omega}$, and minimize the sum of squared Mahalanobis distances $\sum_i d(\hat{\mathbf{e}}_i, \mathbf{e}_i)^2$.

3. Write the parameters of K in vector form as $\mathbf{k} = (f,\ k,\ rf,\ u_0,\ v_0)^\mathsf{T}$ and construct the vector

$$\tilde{\mathbf{k}} = (\mathbf{k}^\mathsf{T},\ \mathbf{l}_1^\mathsf{T},\ \mathbf{l}_2^\mathsf{T},\ \mathbf{l}_2^\mathsf{T}, \mathbf{s}_1^\mathsf{T},\ \mathbf{s}_2^\mathsf{T},\ \mathbf{s}_3^\mathsf{T})^\mathsf{T}$$

where $\mathbf{l}_i,\ i = 1, 2, 3$ is the normalized vanishing line of imaged plane i

$$\mathbf{l}_i = (\frac{\mathbf{l}_{\infty i1}}{\mathbf{l}_{\infty i3}}, \frac{\mathbf{l}_{\infty i2}}{\mathbf{l}_{\infty i3}})^\mathsf{T}$$

and $\mathbf{s}_i,\ i = 1, 2, 3$ encodes the similarity parameters for scale s_i, rotation θ_i and translation $(t_{xi},\ t_{yi})^\mathsf{T}$

$$\mathbf{s}_i = (s_i,\ \theta_i,\ t_{xi},\ t_{yi})^\mathsf{T}$$

$\tilde{\mathbf{k}}$ has 23 parameters, five from $\mathbf{k}$, six from the vanishing lines and 12 from the similarity transformations.

4. Minimize the cost function over $\tilde{\mathbf{k}}$

$$\min_{\tilde{\mathbf{k}}} \sum_i d(\hat{\mathbf{e}}_i, \mathbf{e}_i)^2$$

At each iteration, $\hat{\mathbf{e}}$ is determined by the elements of $\tilde{\mathbf{k}}$ in a straightforward manner: First $\hat{\omega}$ is

computed from $\mathbf{k}$. The imaged circular points lying on $\hat{\omega}$ are computed from the intersection of the vanishing lines $\mathbf{l}_i$ with $\hat{\omega}$ using the Joachimsthal equation described in section 2.2.2. This defines a projective transformation G_i for each square. Vectors $\mathbf{s}_i$ define similarity transformations E_i. The world plane squares defined by the rectifications N_i are then mapped by $\mathsf{G}_i\mathsf{E}_i$ to the image to give $\hat{\mathbf{e}}$.

The camera parameters computed for the image of figure 4.20 appear in table 4.8. The sides of the imaged squares are computed by running a Canny edge detector followed by line fitting and manual selection of the appropriate lines. (It is certainly possible to automate the entire process, but that is beyond the scope of the current discussion.) Corners for the MLE are defined by the intersection of line pairs. Although with six constraints on ω the difference between the linear and MLE results is small, the MLE computation is important in two respects. First it generalizes easily to larger numbers of squares on a plane *and* large numbers of imaged planes (such as multiple views with a fixed camera). Second, the Levenberg-Marquardt computation provides good numerical estimates of the Jacobian, which make first order error analysis simple. The error analysis is described in the following section.

Method	f	k	rf	u_0	v_0
Linear	1143.4	-5.7	1152.1	545.3	371.6
MLE	1143.4	-5.2	1152.5	544.6	371.3

Table 4.8: The camera parameters computed from three imaged squares using the linear and ML methods.

4.5.3 Error propagation

A first order estimate of the error in the estimated camera parameters follows the same procedure as for the MLE of line intersection in section 3.6.2. The minimization algorithm estimates the Jacobian of function g mapping $\hat{\mathbf{e}}$ to $\tilde{\mathbf{k}}$ numerically

$$\mathsf{J}_g^\mathsf{T} = \frac{\partial(\hat{\mathbf{e}} - \mathbf{e})}{\partial\tilde{\mathbf{k}}^\mathsf{T}}$$

and the covariance in the parameter vector $\tilde{\mathbf{k}}$ is given by

$$\Lambda_{\tilde{\mathbf{k}}} = (\mathsf{J}_g^{\mathsf{T}} \Lambda_{\mathbf{e}}^{-1} \mathsf{J}_g)^{+} \tag{4.27}$$

The covariance of K is then a sub-matrix of $\Lambda_{\tilde{\mathbf{k}}}$.

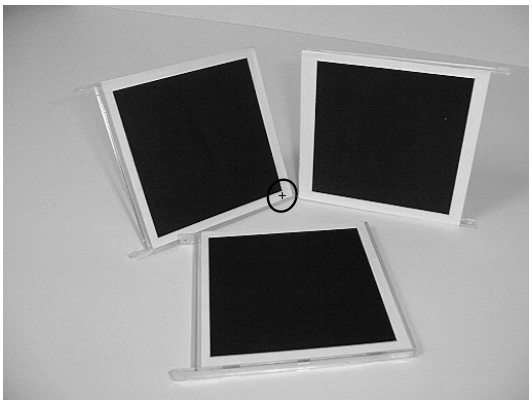
The camera parameters and covariance of the example above assuming a standard deviation of 0.5 pixels on the corners appear in table 4.9. A Monte Carlo simulation has also been performed to test the validity of the results. Synthetic data is obtained from the same example by using the squares defined by $\hat{\mathbf{e}}$. The result of 10 000 iterations adding noise to the corners and computing K appear in table 4.10, and the MLE and its covariance in table 4.11. The results compare favourably, indicating that at these noise levels, the first order error is a reasonable estimate. Plots of the three standard deviation ellipses of the principal point for the Monte Carlo and first order experiments appear in figure 4.23.

	Camera parameters	Covariance matrix				
f	1143.4	134.1190	-13.9892	128.3311	4.7180	1.3870
k	-5.2	-13.9892	41.8636	-13.4386	-10.1093	-19.9487
rf	1152.5	128.3311	-13.4386	160.1897	20.1507	-12.6657
u_0	544.6	4.7180	-10.1093	20.1507	86.9062	-9.3151
v_0	371.3	1.3870	-19.9487	-12.6657	-9.3151	107.1209

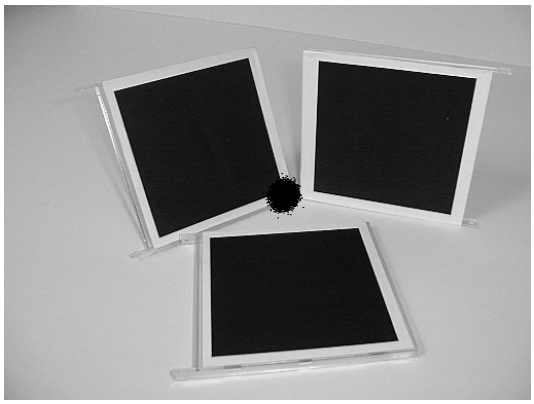
Table 4.9: The camera parameters and first order covariance matrix computed from the ML method on the measured image data using the three square calibration object with a standard deviation of 0.5 pixels on the square corners.

	Camera parameters	Covariance matrix				
f	1143.2	131.7136	-13.4498	125.5439	3.6433	0.1815
k	-5.2	-13.4498	45.6875	-13.3108	-17.3785	-18.1345
rf	1152.3	125.5439	-13.3108	157.7291	21.7178	-13.9876
u_0	544.7	3.6433	-17.3785	21.7178	101.2858	-9.9145
v_0	371.3	0.1815	-18.1345	-13.9876	-9.9145	108.8555

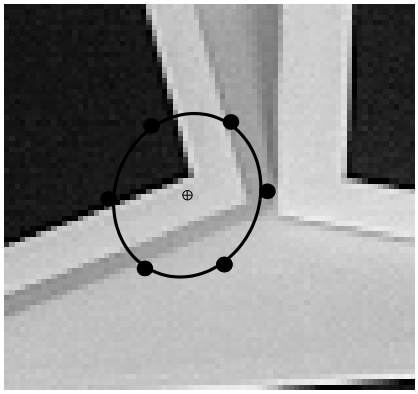
Table 4.10: The camera parameters and first order covariance matrix computed from Monte Carlo simulation using the synthetic three square calibration object with a standard deviation of 0.5 pixels on the square corners in 10 000 iterations.



(a)



(b)



(c)

Figure 4.23: The three standard deviation ellipses of the principal point computed (a) from (4.27). (b) from Monte Carlo simulation. (b) also shows the principal points for all 10 000 iterations. (c) shows the two ellipses from the synthetic data (solid line) and Monte Carlo simulation (filled circles) simultaneously.

	Camera parameters	Covariance matrix				
f	1143.4	134.1189	-13.9892	128.3310	4.7180	1.3870
k	-5.2	-13.9892	41.8636	-13.4386	-10.1093	-19.9487
rf	1152.5	128.3310	-13.4386	160.1896	20.1507	-12.6656
u_0	544.6	4.7180	-10.1093	20.1507	86.9063	-9.3151
v_0	371.3	1.3870	-19.9487	-12.6656	-9.3151	107.1209

Table 4.11: The camera parameters and first order covariance matrix from the ML method using the synthetic three square calibration object with a standard deviation of 0.5 pixels on the square corners.

4.6 Calibration and the rectangle ambiguity

The rectangle ambiguity described in section 3.5.3 occurs with imaged rectangles of unknown aspect ratio. It is possible to rectify the plane so that the orthogonality of the rectangular sides is restored, but an unknown relative scaling in the directions of the sides remains. It was also pointed out that this is a common occurrence in images of building facades, where many rectangular structures exist. This section will show how computing the camera internal parameters provides a method of completing the metric rectification of such a plane.

The key is in the geometric representation of metric information on the plane. When the coordinates of the imaged circular points are known, the metric rectification is fully determined. In the case of an imaged rectangle the vanishing line is known, but there remains a one parameter family of conics dual to the circular points on $\mathbf{l}_\infty$ (see (3.19)). However, if the camera is calibrated, ω is known. Now, since all imaged circular points lie on ω, the plane rectification is defined by the intersection of ω and $\mathbf{l}_\infty$ in the points $\mathbf{I}$ and $\mathbf{J}$. (see figure 4.24).

The unknown components of $\mathbf{I}$ and $\mathbf{J}$ are the affine rectification parameters α and β (defined in chapter 3) . They can be computed by solving

$$\mathbf{I}^\mathsf{T}\omega\mathbf{I} = 0 \tag{4.28}$$

for the unknown α and β. Expanding the expression for $\mathbf{I}$ in terms of α, β and the vanishing line

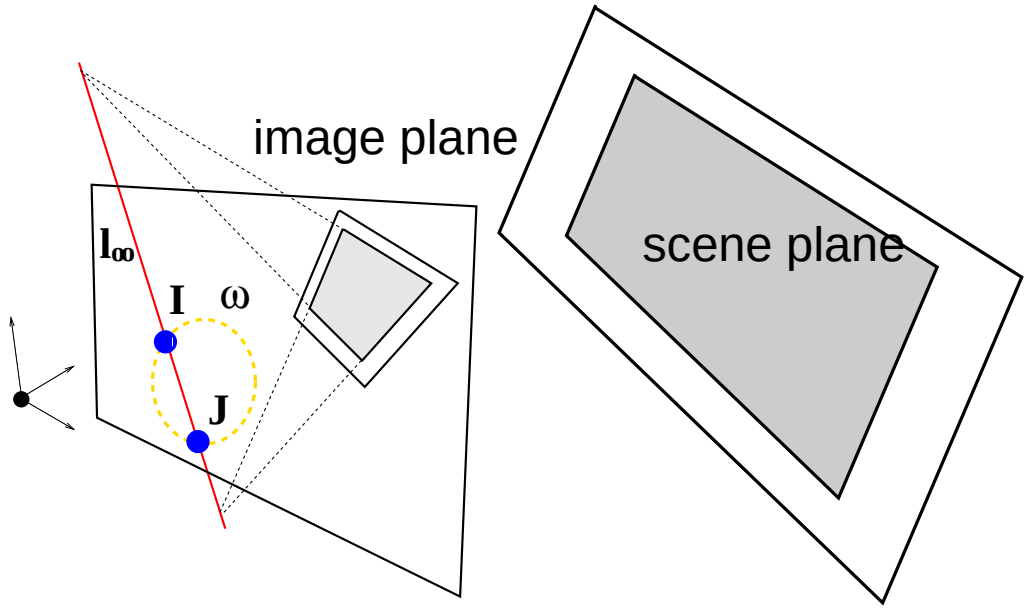


Figure 4.24: The intersection of the vanishing line of a plane and ω fully determines the imaged circular points of the plane, and thus its complete metric rectification.

$\mathbf{l}_\infty$ (as in chapter 3)

$$\mathbf{I} = (I_1,\ I_2,\ I_3)^\intercal = (\alpha - i\beta, 1, \frac{-l_{\infty 2} - \alpha l_{\infty 1} + i l_{\infty 1}\beta}{l_{\infty 3}})^\intercal = (I_1,\ 1,\ \frac{-l_{\infty 2} - l_{\infty 1} I_1}{l_{\infty 3}})^\intercal$$

(4.28) simplifies to a quadratic in I_1

$$\frac{(\omega_1 - 2l_{\infty 1}\omega_4 + l_{\infty 1}^2\omega_6)I_1^2}{l_{\infty 3}^2} + \frac{2(l_{\infty 1}l_{\infty 2}\omega_6 + \omega_2 - l_{\infty 1}\omega_5 - l_{\infty 2}\omega_4)I_1}{l_{\infty 3}^2} + \frac{\omega_3 - 2l_{\infty 2}\omega_5 + l_{\infty 2}^2\omega_6}{l_{\infty 3}^2} = 0 \qquad (4.29)$$

With $l_{\infty 3}$ non-zero and a square pixel camera (4.29) further simplifies to

$$(1 + 2u_0 l_1 + l_1^2(u_0^2 + v_0^2 + f^2))I_1^2 + 2(l_2 u_0 + l_1 v_0 + l_1 l_2(u_0^2 + v_0^2 + f^2))I_1 + 2l_2 v_0 + l_2^2(u_0^2 + v_0^2 + f^2) + 1 = 0 \quad (4.30)$$

The solutions of (4.29) or (4.30) are complex, with α and β given by the real and imaginary parts

respectively.

An earlier version of this approach by Karras[58] uses direct computation of the rectification transformation for a camera with (hard) constraints of zero skew, unit aspect ratio and principal point at the centre of the image. Focal length and the orientation of the world plane relative to the camera follow from two vanishing points.

As an example, consider figure 4.25. The building facade has two dominant directions which are orthogonal, but the rectangle ambiguity exists for the rectification. The vanishing points of the orthogonal directions provide one constraint on ω from (4.2). Assuming the the camera has square pixels, (4.8) provides two constraints. Constraining the principal point to the centre of the image provides two further constraints from (4.10), giving a total of five linear constraints on ω. K can thus be computed and (4.30) applied to compute a metric rectification of the plane appearing in figure 4.25 (b).

A second example, bringing together the rectangle ambiguity, metric plane constraints and calibration degeneracy, appears in figure 4.26. It shows a photograph of the Municipal Theater of Athens, which was destroyed in 1938. The camera is unknown and the image has possibly been cropped, so the principal point cannot be constrained to lie at the image centre. Additionally, there are image parallel lines in three orthogonal directions, but the vertical vanishing point is practically ideal. Thus, the three vanishing points and the square pixel constraints do not fully determine ω. Imaged parallel lines for two orthogonal directions visible on the front facade define the vanishing line of the facade and rectification up to the rectangle ambiguity. The ambiguity cannot be resolved as in the previous example because ω is unknown.

A solution is to complete the metric rectification of the front facade from the ellipses overlayed in figure 4.26 (a). These ellipses are the images of semi-circular arches, and so intersect the vanishing line of the facade in its imaged circular points. Once the metric structure of the facade is known, its imaged circular points constrain the camera. Combining the vanishing point, square pixel and

metric facade plane constraints gives the camera

$$\mathsf{K} = \begin{bmatrix} 1196.86 & 38.76 & 734.18 \\ 0 & 1253.42 & 426.26 \\ 0 & 0 & 1.00 \end{bmatrix}$$

The image is of size 1152×682 pixels, the computed aspect ratio is 1.05 and the angle between the axes is 88.15^o. The square pixel constraints are not perfectly satisfied since the camera is over-constrained. The rectified facade appears in figure 4.26 (b). It is architecturally interesting to note that the assumption that the arches shown in (a) are semi-circular is not consistent with semi-circular arches on the lower tier.

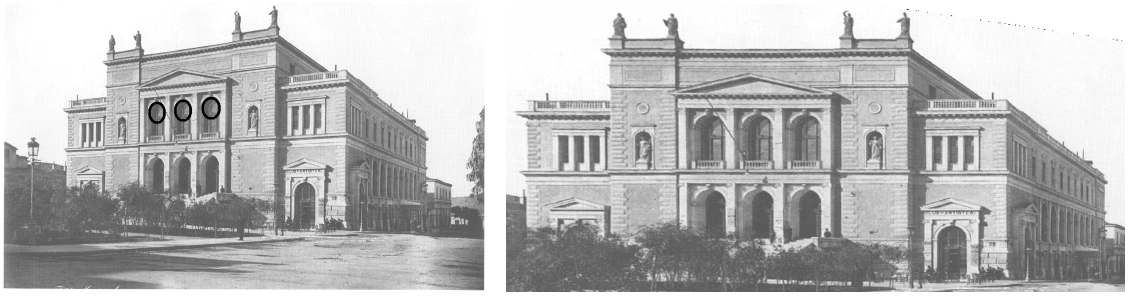
Further examples of calibration and rectification appear in the following section.

(a) (b)

Figure 4.25: Plane rectification via partial internal parameters. (a) Original image. (b) Rectification where the relative scaling of vertical and horizontal directions assumes a square pixel camera with the principal point at the centre of the image. Measurement of the aspect ratio of a window indicates a difference of 3.7% between true and computed values. Note that the two parallel planes, the upper building facade and the lower shop front, are both correctly rectified, but scene planes not parallel to these two are distorted by the rectification homography. This distortion is visible on the area of overhang of the upper facade.

4.7 3D reconstruction from single views

Calibration techniques and the metric plane rectification described in chapter 3 can be combined to produce three dimensional reconstructions of planar objects[66]. In general, the task is to reconstruct a scene from recognizable scene primitives such as lines, planes and spheres by computing



(a) (b)

Figure 4.26: Plane rectification using the properties of circles. (a) The Municipal Theater of Athens, with ellipses overlaid on imaged semi-circular arches. (With thanks to George Karras for providing the image.) (b) The rectified front facade.

their spatial layout. The main application is to buildings, mostly planar structures where there are plentiful constraints to be had from parallelism and orthogonality. Reconstruction is fundamentally model based – the architectural scene is described by a cuboid model, and calibration and rectification allow the parameters of the cuboid to be computed. This section describes such reconstruction from single views and sets of minimally overlapping views. Minimally overlapping views are cases where multiple view methods are inapplicable because insufficient points are visible in more than one view, but images have sufficient scene structure in common to stitch individual single view reconstructions together.

Building architectural models in this manner is similar to the photogrammetric techniques of Debevec *et al.*[25]. However, in the Debevec system scene measurements are required to position the cameras, and the camera internal calibration must be known. These are not needed here. The use of several vanishing points for scene modeling from single images also extends the work of Horry *et al.*[54], where a single vanishing point is used. A single view reconstruction system using the same approach as described above for calibration has been developed by Sturm and Maybank[98]. Their system allows a user to define points and parallel lines as well as orthogonal directions. Points and planes are then reconstructed simultaneously using a constrained optimization. Single view reconstruction using a metric ground plane and a vanishing points off the plane has been demonstrated by Criminisi *et al*[21]. The distances of points in the direction of the vanishing point are computed relative to a reference, so the vertical vanishing point and one height measurement are sufficient to

reconstruct a building.

The single view techniques are complementary to reconstruction methods applicable to multiple images, such as an image sequence acquired by a video camera when walking around a building[4, 103, 112]. More will said about multiple view techniques in the next chapter. It must, however, be noted that single view reconstructions are *not* complete 3D reconstructions in the same sense as stereo or multiple view reconstructions. A single view reconstruction only allows world measurements to be made on particular sets of planes or in specific directions. Measurement of properties such as the 3D distance between two general points is not possible.

4.7.1 Single view reconstructions of buildings

The general approach to reconstructing buildings is to model the building as a cuboid, with three orthogonal planes visible in the image. These planes provide parallel lines in three orthogonal directions from the facade outlines and rectangular features such as doors and windows. The aim is to measure the dimensions of the cuboid and adjacent planes forming part of the structure, as illustrated in figure 4.27.

There are two components to the reconstruction process. First, the camera calibration and metric structure of each plane is computed. The second stage involves stitching the individually rectified planes together, accounting for the arbitrary scale factors in the rectification of each plane and the relative orientation of planes not known to be orthogonal.

The first part, the calibration and rectification process, comprises three steps:

1. Calibrate a square pixel camera from three vanishing points of orthogonal directions, also using the known principal point constraints if possible.

2. Compute the vanishing line of each plane in the image. Again, this is usually done from the vanishing points of imaged parallel lines on each plane. The vanishing lines define affine rectification of each of the planes.

3. Metric rectify each plane by computing the imaged circular points from the intersection of the vanishing line and ω using (4.30).

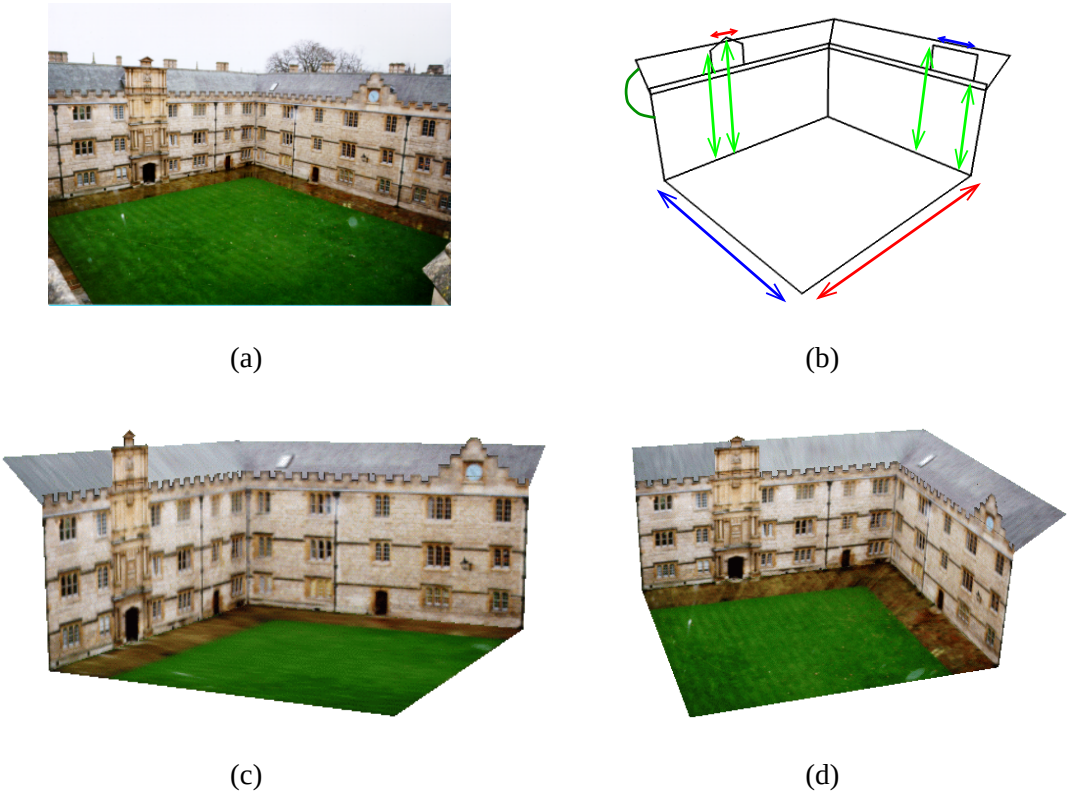


Figure 4.27: 3D reconstruction from a single image. (a) Fellows quad, Merton College, Oxford. (b) The dimensions (up to a global scale) required to compute a reconstruction. (c) and (d) Novel views of the textured 3D model created from the single image in (a).

Once all building planes are rectified, model construction proceeds by assigning them appropriate Euclidean co-ordinates. The first consideration in this is the relative scale of planes. Each metric rectified plane has an arbitrary global scale factor (see figure 4.28), so the planes must be scaled for consistency. The relative scaling of a pair of planes can be computed from any two points common to both planes by finding the two lengths between the common points defined by the two rectification homographies of the planes. The Euclidean dimensions of one plane can then be scaled by the ratio of these lengths to make it consistent with the other plane.

The steps described thus far are sufficient to reconstruct the orthogonal planes defining the cuboid. Taking the left facade in figure 4.27(a) as reference, for example, its correctly proportioned width and height are determined by rectification. The right facade and ground planes are orthog-

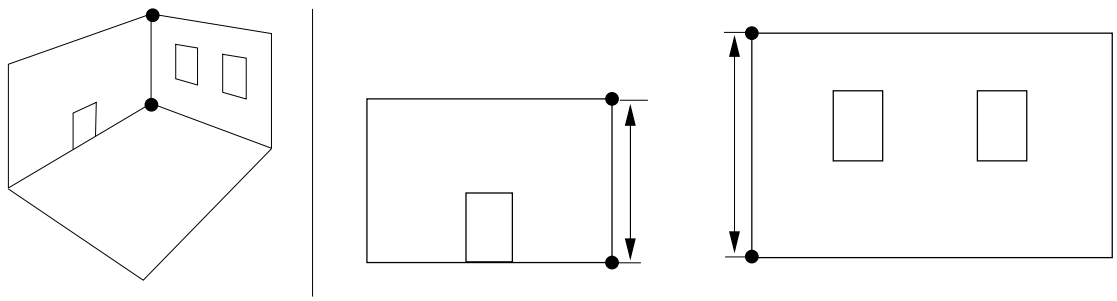


Figure 4.28: Two points common to a pair of planes have different lengths in the two rectifications because each point pair is known up to an arbitrary scale factor.

onal to the reference, although there are insufficient scene constraints to rectify the ground plane independently. However, scaling of the rectified right plane from the points common to the planes where the facades meet completes the cuboid model. The dimensions of the cuboid are now known, and the ground plane rectification homography can be computed from these dimensions.

The final aspect of the reconstruction is to determine the relative orientation of planes that are not obviously orthogonal. Computing the relative orientation of a pair of planes requires a calibrated camera and the vanishing lines of both planes, as will now be explained.

The vanishing line and the camera together determine the orientation of the world plane relative to the camera[15] as follows. The normal to a plane $\boldsymbol{\pi}$ in camera centred co-ordinates is

$$\mathbf{n}_{\boldsymbol{\pi}} = \mathsf{K}^{\mathsf{T}} \mathsf{l}_{\infty}$$

This result follows from considering a 3D plane $\boldsymbol{\pi} = (\pi_1,\ \pi_2,\ \pi_3,\ \pi_4)^{\mathsf{T}}$. The orientation of the plane is described by the normal vector $\mathbf{n}_{\boldsymbol{\pi}} = (\pi_1,\ \pi_2,\ \pi_3)^{\mathsf{T}}$. Plane $\boldsymbol{\pi}$ intersects the plane at infinity $\boldsymbol{\pi}_{\infty} = (0,\ 0,\ 0,\ 1)^{\mathsf{T}}$ in the line $\mathbf{n}_{\boldsymbol{\pi}}$. This line is imaged as the vanishing line of the plane $\mathsf{l}_{\infty} = \mathsf{K}^{-\mathsf{T}} \mathbf{n}_{\boldsymbol{\pi}}$. The angle θ between two planes then follows from the inner product of the two

plane normals:

$$\begin{aligned}\cos(\theta) &= \frac{\mathbf{n}_{\boldsymbol{\pi}1}^{\mathsf{T}}\mathbf{n}_{\boldsymbol{\pi}2}}{\|\mathbf{n}_{\boldsymbol{\pi}1}\|\,\|\mathbf{n}_{\boldsymbol{\pi}2}\|} \\ &= \frac{\mathbf{l}_{\infty 1}^{\mathsf{T}}\omega^{-1}\mathbf{l}_{\infty 2}}{(\mathbf{l}_{\infty 1}^{\mathsf{T}}\omega^{-1}\mathbf{l}_{\infty 1})^{1/2}(\mathbf{l}_{\infty 2}^{\mathsf{T}}\omega^{-1}\mathbf{l}_{\infty 2})^{1/2}}\end{aligned} \tag{4.31}$$

The roof planes in figure 4.27(a) are reconstructed using (4.31). The vanishing lines are obtained the regular pattern of tiles, using an automatic method developed by Schaffalitzky and Zisserman[90]. In this method, the geometric grouping of imaged equally spaced parallel lines defines the vanishing line by invariance of the cross ratio of the line pencil[1].

A second example appears in figure 4.29. Four planes are clearly visible in the scene, three of which are orthogonal. In addition, sufficient parallel line sets are available to determine the camera internal parameters and vanishing lines for each plane. All planes can thus be rectified, and the relative orientation of the non-orthogonal plane determined.

4.7.2 Multiple single views

It is often difficult to obtain sequences of images of buildings that are suitable for multiple view reconstruction techniques. Occlusions and viewpoint restrictions due to the presence of other buildings can cause severe difficulties. With only a minimal proportion of the scene in common, however, it is possible to stitch together a number of single view reconstructions to obtain a complete model of a building.

The requirement for areas of overlap are the same as for planes in a single view. To join the reconstruction of planes from parts visible in two views there must be common world points in both views. This allows the relative scale of the two rectifications to be computed.

Two examples follow, both reconstructions of Oxford college quads. The first is Fellows quad, Merton College, the same scene that appears in figure 4.27. The second is in Oriel College, a model created to advertise a student ball. The reconstruction is augmented with additional structures and a rather fanciful fire (all at the request of the students involved). The augmentation process exploits

[1] with thanks to Frederick Schaffalitzky for computing the roof vanishing lines.

(a) (b) (c)

Figure 4.29: Single view reconstruction. (a) The Radcliffe Observatory, Oxford. (b) and (c) Views of the textured reconstruction. The angle between the non-orthogonal plane and the facade planes has been computed from (4.31) to be 46°. The true angle is 45°.

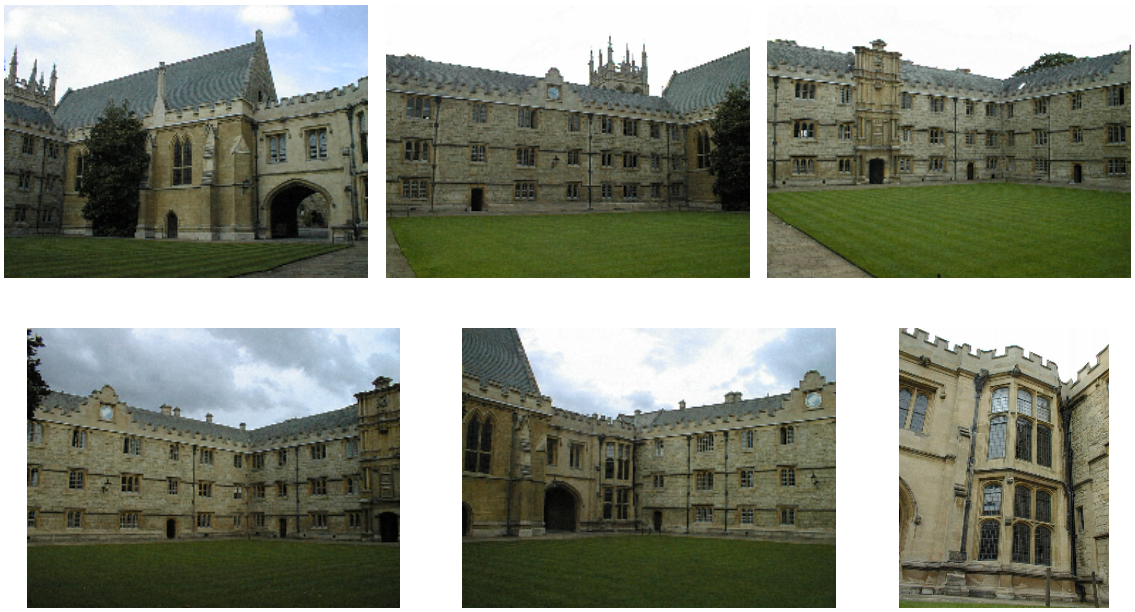
the power of VRML to create the additions with textured images and appropriate lighting.

4.8 Summary

This chapter has examined single view calibration of a camera from scene and internal parameter constraints. Scene constraints are obtained from the vanishing points of orthogonal directions and metric scene planes, exploiting the geometric properties common to built environments. Internal parameter constraints are written from knowledge of some properties of the camera.

Particular attention has been paid to the case of a square pixel camera, which provides calibration constraints identical in form to a metric scene plane. The common case of such a camera viewing a scene where the vanishing points of three orthogonal directions can be computed has been analysed, detailing the degenerate conditions under which this configuration does not fully constrain the camera. A first order error analysis of the square pixel and three vanishing point configuration has been presented and verified with Monte Carlo simulation. The error analysis results are consis-

Source images of the quad.



Three views of the reconstruction.

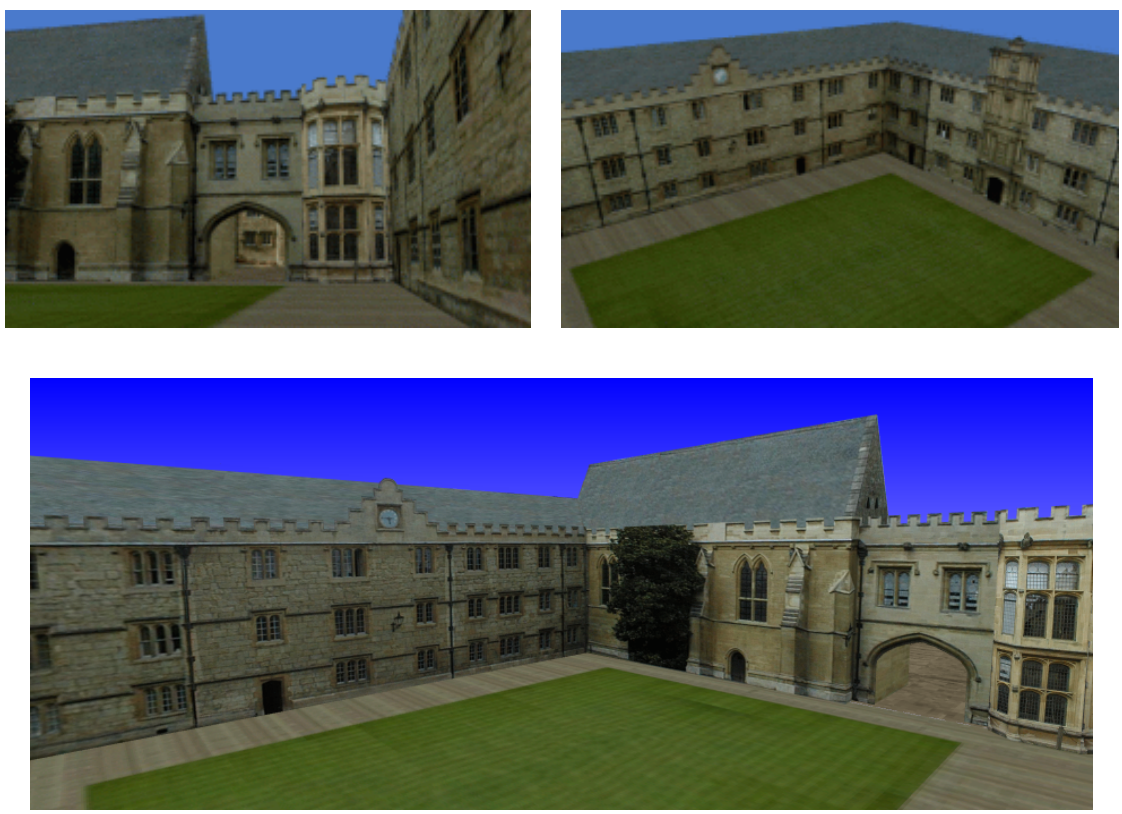


Figure 4.30: Source images and views of the multiple single view reconstruction of Fellows quad, Merton College.

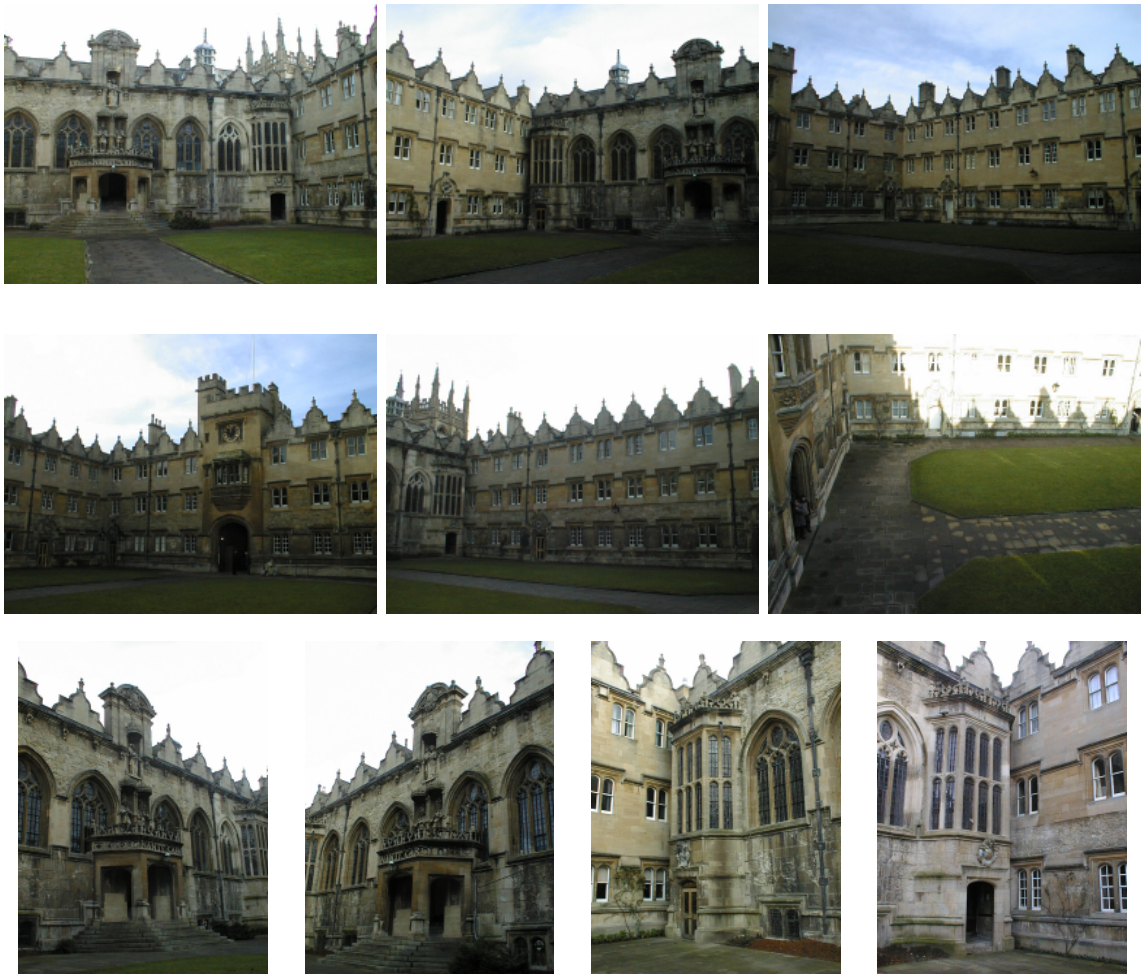


Figure 4.31: Source images of Oriel College, Oxford, for the augmented reconstruction.

tent with the degenerate conditions, and the effect of introducing a known principal point constraint has been examined.

The metric plane calibration constraints have been exploited to create a calibration object consisting of squares, and linear and Maximum Likelihood techniques of computing the camera described. A first order error analysis for the latter case has also been presented and verified.

Camera calibration has been applied to single view reconstruction of architectural scenes. The metric rectification described in chapter 3 and single view calibration have been combined to produce accurate and realistic models of buildings from single views and sets of minimally overlapping views.

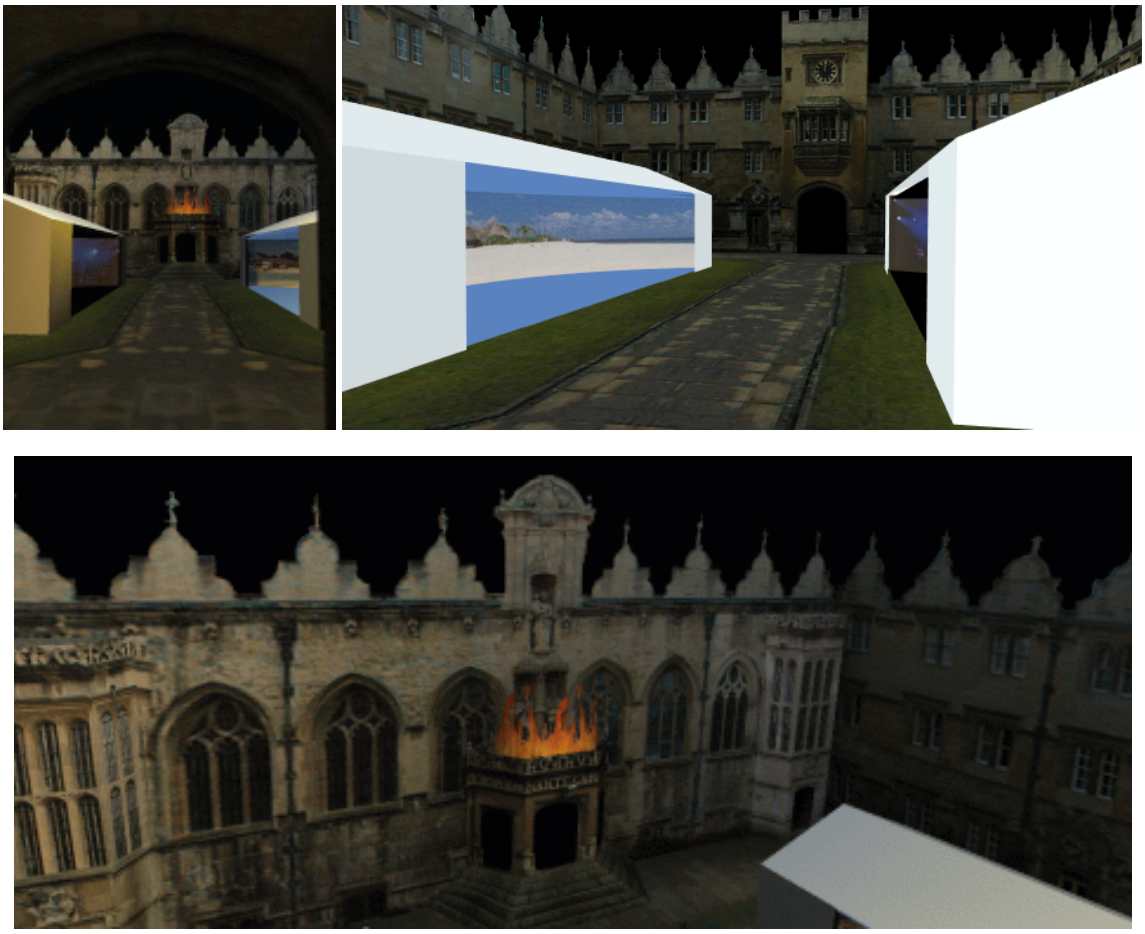


Figure 4.32: Views of the augmented reconstruction of Oriel College.

The following chapter extends square pixel camera calibration using vanishing points to two views. The constraints valid for a single view apply to each camera, and can also be transferred between views, providing additional constraints from motion. It will be shown how, even if a degenerate condition in each view is present, calibration and metric 3D reconstruction are still possible.

Chapter 5

Calibration and Reconstruction in Two Views

5.1 Introduction

The task directly addressed in this chapter is as follows. Given two views of a scene containing a cuboid object with corresponding point sets, compute a metric reconstruction. As an example, consider figure 5.1, two views of a building, and figure 5.2, where the textured, reconstructed building created with the methods described in this chapter is seen from novel viewpoints. The cuboid object defines the vanishing points of three orthogonal directions, as in the case with single view reconstruction. There is an important difference, however, between a stereo reconstruction of such an object and a single view reconstruction. A stereo reconstruction allows true 3D measurement of the scene, since any point that is visible in both views may be back-projected, rather than the restricted metric spaces obtained from a single view.

The approach taken here assumes that the epipolar geometry of the view pair, and thus a projective reconstruction, is available. A metric reconstruction is obtained from the projective by the stratified scheme described in section 2.3.3: the plane at infinity and the image of the absolute conic are computed to yield a metric reconstruction.

The chapter proceeds with a description of the constraints on the plane at infinity and IAC

Figure 5.1: Two views of a building.

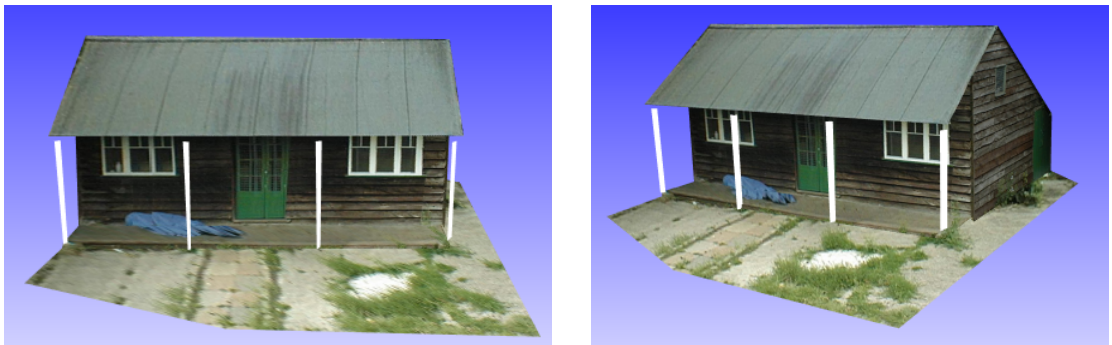


Figure 5.2: Two novel views of the reconstructed building.

obtained from two views of a cuboid object. The case that both cameras have square pixels is developed in some detail, and its implementation in a two view metrology tool is described in section 5.3. Section 5.4 returns to theoretical issues, specifically the additional constraints available when the two cameras have identical internal parameters.

5.2 Calibration constraints

The task of calibrating a pair of cameras is approached here in the much the same manner as single view calibration. The vanishing points of three orthogonal directions constrain the image of the absolute conic in either view. Further constraints are obtained when the cameras have square pixels. Important additional constraints exist, however, because the vanishing points are computed in both views. The three matched vanishing points are the images of points on π_∞, so the plane can be computed and the structure affine rectified from (2.10). The infinite homography, H_∞, which maps

projections of all ideal points between views is also computed. This allows constraints such as the known image plane circular points (for square pixel cameras) and the invariance of ω (for two cameras with identical parameters) to be transferred between views. It is thus the motion between cameras that introduces further constraints on the cameras in either view.

A similar approach, but without reliance on computing epipolar geometry, is reported by Cipolla and Boyer[12]. Their method relies on first computing the calibrations in each view independently from three vanishing points assuming zero skew and known principal point. The rotation between views is computed as the transformation between the vanishing point sets after correction for the internal parameters. Translation follows from the observation that the translation vector and the rays through the camera centre defined by any matched point are coplanar. While no direct comparison of the two approaches has been made, it is likely that using accurately determined epipolar geometry will result in a more accurate calibration since it uses motion constraints in addition to structural constraints. However, in cases where it is difficult to accurate estimate the epipolar geometry, their approach is probably preferable.

5.2.1 Vanishing points in two views

An ideal 3D point projects to a vanishing point in each of two views (see figure 5.3). The vanishing points can be computed as before, from imaged parallel lines, such as those shown in the two views of figure 5.4. Note that it is the direction of the lines that determines the vanishing point, so the same lines need not be visible in both views.

The three matched vanishing points define both the plane at infinity and the infinite homography if the epipolar geometry is known. Given an initial projective reconstruction and back-projecting the vanishing points $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_2$ in the first view and $\mathbf{v}'_1$, $\mathbf{v}'_2$ and $\mathbf{v}'_3$ in the second gives three 3D points $\mathbf{V}_1$, $\mathbf{V}_2$ and $\mathbf{V}_3$ on $\boldsymbol{\pi}_\infty$. The homogeneous plane representation for $\boldsymbol{\pi}_\infty$ follows from a null space computation described in section 2.2.3 and H_∞ from (2.14). Affine structure and projections are now defined. The projective model of figure 2.7, for example, can be affine rectified, as in figure 5.5.

The pairs of vanishing points of orthogonal directions constrain the internal parameters of each

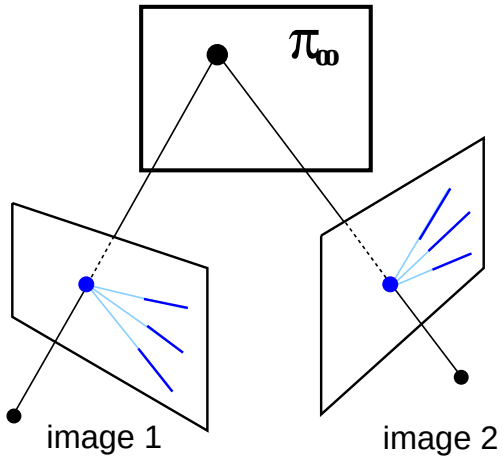


Figure 5.3: An ideal point is projected to a vanishing point in each of two views. The vanishing points can be computed from imaged parallel lines in each view (not necessarily the same lines).

Figure 5.4: Imaged parallel line segments shown in two views of a building define three vanishing points in each view.

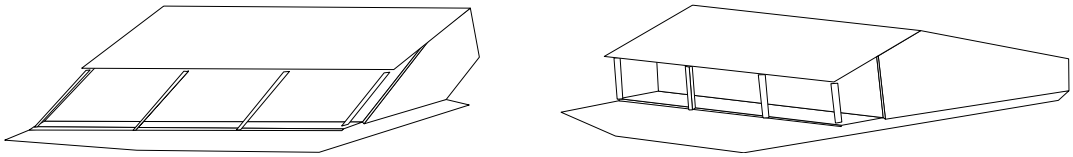


Figure 5.5: Two views of an affine reconstruction of the building of figure 5.1. Note that affine properties of the structure are correctly represented (parallel world lines are parallel in the reconstruction), but metric properties remain distorted. The model is shown in orthographic projection (rather than perspective) to make this clearer.

camera from the conjugacy constraint (4.2). In the vector notation of (4.4), these constraints are $\boldsymbol{\kappa}_{\mathbf{v}_1\mathbf{v}_2}$, $\boldsymbol{\kappa}_{\mathbf{v}_1\mathbf{v}_3}$, $\boldsymbol{\kappa}_{\mathbf{v}_2\mathbf{v}_3}$ in the first view. There are thus three constraints on the five parameters of ω, leaving a two parameter ambiguity.

Just as in the single view case, further constraints can be written on each camera if it has square pixels. It will be shown now how the constraint can be transferred between views.

5.2.2 Square pixel cameras and the motion constraint

The constraint that the first camera has square pixels, (4.8), provides two additional constraints on ω since image plane circular points $\mathbf{I}$ and $\mathbf{J}$ lie on ω. There are then five constraints on ω, which may be reduced in degenerate cases to four (section 4.3.2). Now, if the second camera has square pixels the second image plane circular points are $(1,\ \pm i,\ 0)^\mathsf{T}$ in image co-ordinates. Additionally, H_∞ is known, so these image plane circular points and the ideal line can be transferred to the first view (figure 5.6), giving

$$\mathbf{I}^* = \mathsf{H}_\infty^{-1}\,(1, i, 0)^\mathsf{T} \quad \text{and} \quad \mathbf{J}^* = \mathsf{H}_\infty^{-1}\,(1, -i, 0)^\mathsf{T}$$

with $\mathbf{I}^*$ and $\mathbf{J}^*$ the second image plane circular points as seen by the first camera.

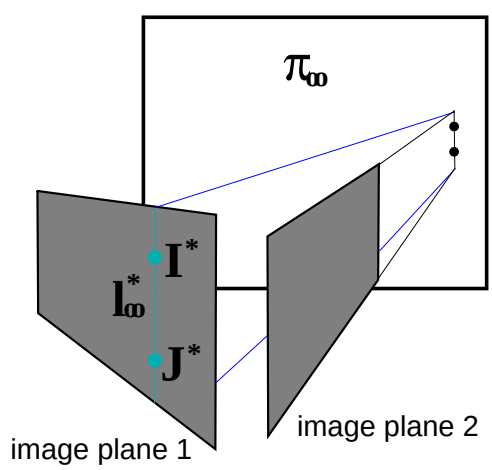


Figure 5.6: An ideal point is projected to a vanishing point in the second view, and mapped to the first by H_∞. This applies to the image plane circular points, known if the second camera has square pixels, which are mapped to $\mathbf{I}^*$ and $\mathbf{J}^*$ in the first view.

$\mathbf{I}^*$ and $\mathbf{J}^*$ are projections of points on Ω_∞, and thus lie on ω in the first view. Two additional constraints on ω are available from $\mathbf{I}^{*\mathsf{T}}\omega\mathbf{I}^* = 0$ and $\mathbf{J}^{*\mathsf{T}}\omega\mathbf{J}^* = 0$. The two constraint coefficient vectors $\boldsymbol{\kappa}_{\mathbf{I}^*}$ and $\boldsymbol{\kappa}_{\mathbf{J}^*}$ follow from expanding $\mathbf{I}^{*\mathsf{T}}\omega\mathbf{I}^* = 0$ and taking real and imaginary parts.

There are thus a total of seven constraints on ω to be found from two square pixel cameras. A construction appears in figure 5.7.

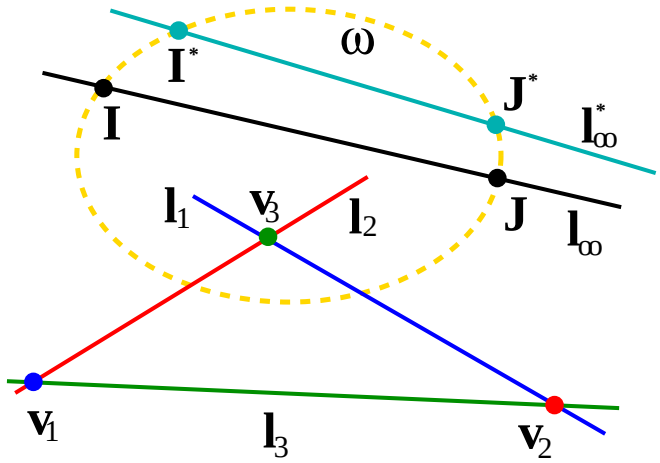


Figure 5.7: Seven constraints on ω, three from the conjugacy of the vanishing points of three orthogonal directions, two from the image circular points, and two from the transferred circular points.

5.2.3 Square pixel degeneracies

Degenerate conditions occur, and can be treated similarly to the single view case described in section 4.3.2. The vanishing points of three orthogonal directions determine a two parameter family of conics $\mathsf{C}(\mu, \nu, \eta)$ parametrized by lines forming the sides of the self-polar triangle (4.13). Constraints on the remaining two parameters can be obtained from the imaged circular points of the first image plane and the transferred circular points of the second image plane as described above. The transfer, by H_∞, depends on the relative orientation of the two cameras, so the rotation between views thus effects the number of constraints on ω.

The degenerate condition that the circular points on a line through a vertex of the self-polar triangle provide only a single constraint on $\mathsf{C}(\mu, \nu, \eta)$ applies. Intersections of the ideal line of the first image plane $\mathbf{l}_\infty$ and the transferred ideal line of the second image plane $\mathbf{l}^*_\infty$ with vertices of the

self-polar triangle thus induce degeneracies. If only one of the image plane ideal lines intersects a vertex of the triangle there are six independent constraints on ω. This might occur as in the single view case, with one or two vanishing point ideal, or might be the result of one or two finite vanishing points lying on $\mathbf{l}_\infty^*$. The latter implies that, since incidence is preserved under homography, the relevant vanishing points are ideal in the second image. To clarify, say $\mathbf{l}_\infty^*$ intersects vertex $\mathbf{v}$ in the first view, it follows that $\mathbf{v}' = \mathsf{H}_\infty \mathbf{v}$ is ideal in the second view:

$$\mathbf{v}^\mathsf{T}\mathbf{l}_\infty^* = \mathbf{v}^\mathsf{T}\mathsf{H}_\infty^\mathsf{T}(0,0,1)^\mathsf{T} = \mathbf{v}'^\mathsf{T}(0,0,1)^\mathsf{T} = 0$$

If $\mathbf{l}_\infty$ and $\mathbf{l}_\infty^*$ *both* intersect a vertex, there is only a single constraint from each pair of circular points, giving a total of five constraints. Note that if there is no rotation between cameras $\mathbf{l}_\infty$ and $\mathbf{l}_\infty^*$ coincide, and the situation is identical to the single view case.

It is common in practice that $\mathbf{l}_\infty$ and $\mathbf{l}_\infty^*$ both intersect a vertex of the self-polar triangle. The two views in figure 5.1, for example, are both imaged with the camera plane vertical (the axis of rotation between views is vertical). The vertical vanishing point is thus ideal in both views and intersects both $\mathbf{l}_\infty$ and $\mathbf{l}_\infty^*$ in the first view (see figure 5.8 for a construction). There are a total of five constraints on ω in either view.

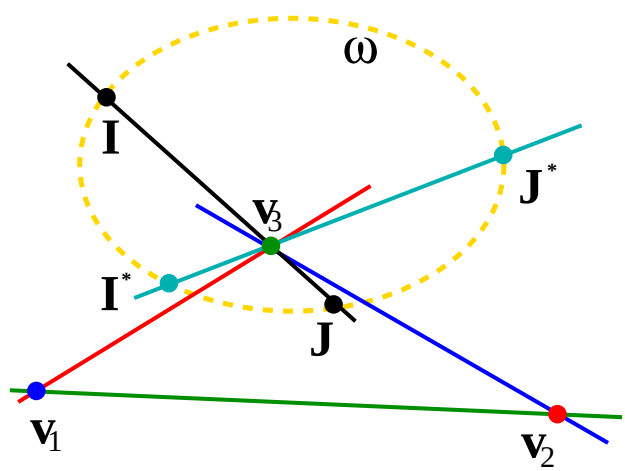


Figure 5.8: Degenerate condition. Five constraints on ω, three from the conjugacy of the vanishing points of three orthogonal directions, one each from the first view imaged circular points and the transferred second view circular points.

5.2.4 Metric structure from calibration

The calibration constraints obtained from three vanishing points and square pixels in both views allows computation of the internal parameters of both cameras and metric rectification from (2.11). The example of figure 5.1, which was seen to be degenerate, still provides five constraints on ω. The computed cameras for that example are

$$\mathsf{K}_1 = \begin{bmatrix} 1089.71 & 1.01 & 442.70 \\ 0 & 1089.66 & 380.32 \\ 0 & 0 & 1.00 \end{bmatrix} \quad \text{and} \quad \mathsf{K}_2 = \begin{bmatrix} 1015.97 & -0.39 & 493.78 \\ 0 & 1016.06 & 450.68 \\ 0 & 0 & 1.00 \end{bmatrix}$$

and the metric rectified structure is shown in figure 5.9. The textured structure appeared earlier, in figure 5.2. A comprehensive description of the rectification process detailing the computational steps at each stage follows.

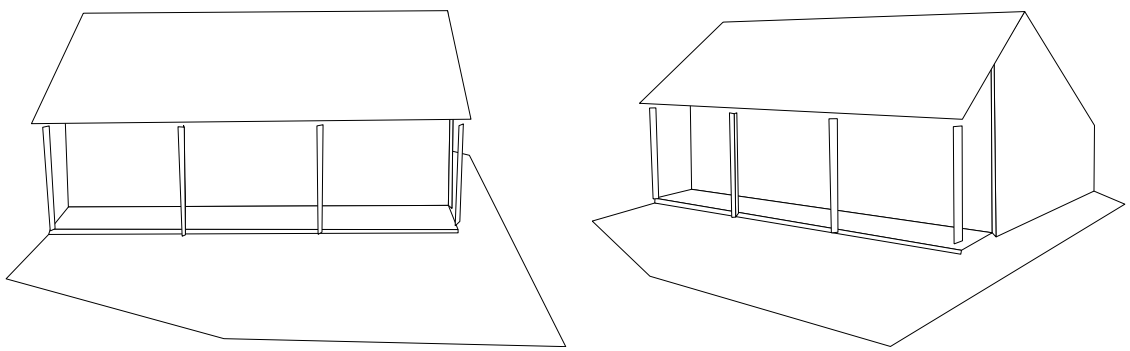


Figure 5.9: Two views of the metric reconstruction of the building of figure 5.1. The metric reconstruction is obtained from the affine by calibrating the cameras in each view using the vanishing points of three orthogonal directions and the constraints that both cameras have square pixels.

5.3 Implementation: 3D metrology from two views

This section describes a 3D measurement system based on two square pixel views of the vanishing points of three orthogonal directions. It is thus suited to images of built environments, mainly indoors or outdoor building scenes where there are many rectangular features. The system requires the user to manually select matched points and sets of parallel lines in the images. These are used

to compute an initial projective reconstruction and its metric rectification. Line segments of known length can then be defined to calibrate the overall scale of the system, finally allowing the Euclidean distance between any two pairs of matched points to be computed. The details of the algorithm implemented by the system (and used to produce the results seen in the previous sections) will now be described. The system is implemented using the TargetJr libraries[17].

5.3.1 Line segments and matched points

The first computation performed by the system is to find line segments along intensity edges in the image. The line segments are important for two reasons. Firstly, they are used to define points in the images by line intersection. This is useful for points at corners of structural features such as walls and doorways, and is more accurate than point selection with a mouse. Second, the line segments defining imaged parallel lines are required for the calibration stage.

Line segments are detected by: Canny edge detection at sub-pixel accuracy[10]; edge linking; segmentation of the edgel chain at high curvature points; and finally, straight line fitting by orthogonal regression to the resulting chain segments. Figure 5.10 shows the line segments computed for a pair of building images, with a feature point defined by line intersection isolated in figure 5.11.

Figure 5.10: Line segments computed for an image pair of a building in Wadham college.

The epipolar geometry of the image pair must be computed to define an initial projective recon-

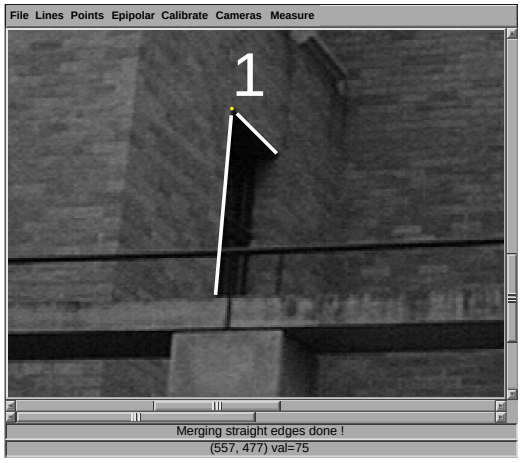


Figure 5.11: Defining a point by line intersection. A pair of lines accurately defines the corner of a doorway.

struction of the scene. The fundamental matrix is computed in a two stage process, first computing an *initial* estimate from a user supplied set of point matches. This estimate is subsequently refined using a guided search process.

Experience indicates that in a typical building scene 25 to 30 points, distributed in the images, provides a reasonable initial estimate of F. Two or more lines intersecting in a feature point are manually selected, and the intersection computed using the MLE method from section 3.6. Additional points can be defined by direct mouse clicking. This is less accurate than line intersection, and is often not needed in building scenes. Figure 5.12 illustrates a set of matched points obtained by line intersection alone.

The initial estimate of F is computed directly from the manual correspondences using the method of Luong and Faugeras[70]. Its accuracy, however, can be significantly improved by a denser set of correspondences. Typically, hundreds of matches are required to provide a sufficiently accurate fundamental matrix. However, since an initial estimate is available, a search for further correspondences is guided by the estimated epipolar geometry. The refinement process begins with computation of a set of interest points in the images, identified by Harris corners[42], a typical step in automatic computation of epipolar geometry used in long sequence algorithms such as those of Fitzgibbon and Zisserman[35]. Robust correspondence matching follows using the random sam-

Figure 5.12: A set of user defined matched points for initial computation of the fundamental matrix.

pling and consensus (RANSAC) approach[104]. A minimal set of seven points in the first view is chosen at random, with matches in the second view guided by the initial estimate of the fundamental matrix. A new fundamental matrix is computed and its *support* evaluated by calculating the error in epipolar matches for the remainder of the data. This process is repeated until a fundamental matrix with the support of the majority of the corners is obtained.

A pair of projective cameras are computed from the final epipolar geometry in the form of (2.5) and (2.6), paving the way for stratified calibration. Note also that Hartley-Sturm[48] correction of the image points is performed in subsequent calculations: image point pairs are adjusted to be consistent with epipolar geometry. Specifically, given a pair of matching points $\mathbf{x}$ and $\mathbf{x}'$, $\mathbf{x}'^{\mathsf{T}}\mathsf{F}\mathbf{x} \neq 0$. Hartley-Sturm correction results in a pair of points $\hat{\mathbf{x}}$ and $\hat{\mathbf{x}}'$ such that $\hat{\mathbf{x}}'^{\mathsf{T}}\mathsf{F}\hat{\mathbf{x}} = 0$ and the distances $d(\mathbf{x}, \hat{\mathbf{x}})$ and $d(\mathbf{x}', \hat{\mathbf{x}}')$ are minimized.

5.3.2 Calibration

The calibration stage requires the user to select three orthogonal sets of imaged parallel lines in both views (figure 5.13). The three matched vanishing points then define the infinite homography and constrain the internal camera parameters.

The vanishing points are computed from imaged parallel line segments which, due to noise, do

Figure 5.13: User defined sets of parallel lines in two views

not intersect in a unique point. As in the single view case, a maximum likelihood estimate (MLE) of each vanishing point is computed, but taking into account both views in an iterative minimization process.

Initially, each vanishing point is found independently in each view as in section 3.6. A corresponding point pair is Hartley-Sturm corrected and back-projected to give a 3D point $\mathbf{V}$. The MLE is then obtained by varying $\boldsymbol{V}$, which projects to the vanishing points $\mathbf{v}$ and $\mathbf{v}'$ in the first and second views respectively [32] (see figure 5.14). The ML estimate of $\boldsymbol{V}$ is determined by a simple extension of the single view vanishing point computation taking into account errors measured in both views: by minimizing a cost function of the sum of squared orthogonal distances from the end points of the imaged parallel lines to lines through $\mathbf{v}$ in the first view *and* through $\mathbf{v}'$ in the second, using the Levenberg-Marquardt algorithm.

Three matched vanishing points define $\boldsymbol{\pi}_\infty$, H_∞ and affine structure. Metric structure follows from the internal parameters of the camera in the first view. This is computed using the three conjugacy constraints from the triad of vanishing points and the square pixel constraints from both views.

The final calibration step required is to define the global scale of the reconstruction. The user selects the matched endpoints of one or more line segments of known 3D length (figure 5.15). A

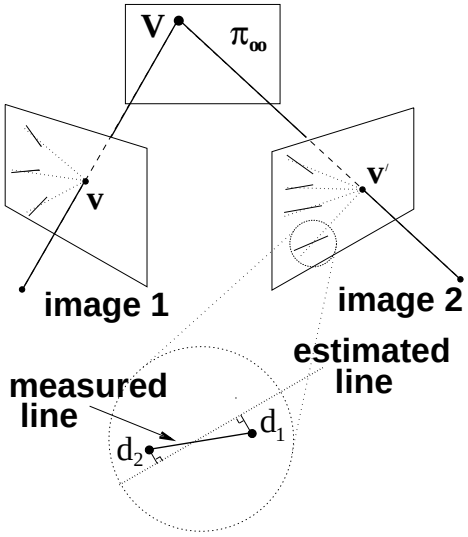


Figure 5.14: MLE Computation of π_{∞} from vanishing points.

global isotropic scaling s is computed from the ratio between the metric reconstructions of these lengths and the true lengths.

Figure 5.15: User defined line segments of known length.

With a calibrated system, the user is able to measure world distances between selected points, as in figure 5.16.

Figure 5.16: 3D measurement of distances between selected points.

5.4 Cameras with identical internal parameters

Two views from a camera with fixed internal parameters (or two identical cameras) provide constraints on ω when H_∞ is known. The constraints arise from the observation of Luong and Viéville[71] that the IAC is unchanged under transformation by H_∞. That is, with $\mathsf{K}' = \mathsf{K}$ and $\omega' = \omega$,

$$\omega = \mathsf{H}_\infty^\mathsf{T} \omega \mathsf{H}_\infty \tag{5.1}$$

The initial formulation of the constraints follows from expanding (5.1) in each of the six unique elements of ω. The resulting six constraints are not independent, but have a total rank of four. The following sections describe the four constraints geometrically and examine the degeneracies that occur when they are applied in the context of a triad of vanishing points visible in both views.

5.4.1 Constraints from identical cameras

The four calibration constraints arising from a pair of identical cameras can be identified explicitly in terms of the imaged axis of the rotation between cameras and the plane orthogonal to this axis [117]. The vanishing point of the rotation axis, $\mathbf{a}$, and the imaged circular points of the plane

orthogonal to the axis, $\mathbf{I_a}$ and $\mathbf{J_a}$ are the fixed points of H_∞. They are thus the eigenvectors of H_∞.

The genesis of this computation is fairly straightforward [116]. Under 3D rotation, there are three fixed points on the plane at infinity corresponding to the three eigenvectors of the rotation matrix. Its one real eigenvector, corresponding to the unity eigenvalue, is the direction of the axis of rotation. The complex conjugate pair of eigenvectors, with eigenvalues $e^{\pm\theta}$, are the circular points of the plane orthogonal to the rotation axis. Geometrically, it is clear that the projection of the direction of the axis of rotation, the vanishing point of the axis, is invariant to the rotation. The planes orthogonal to this axis experiences a 2D plane rotation, and have a pair of fixed points on the plane at infinity, the similarity invariant circular point pair. Now, comparing Euclidean and affine cameras, it can be shown that

$$\mathsf{H}_\infty = \mathsf{K}\mathsf{R}_\mathsf{c}\mathsf{K}^{-1} \quad \text{if} \quad \mathsf{K}' = \mathsf{K}$$

The axis of rotation can be written as a homogeneous ideal point $(\mathbf{r}^\mathsf{T},\ 0)^\mathsf{T}$, which is imaged as $\mathbf{a} = \mathsf{K}(\mathsf{R}_\mathsf{c}|\ \mathbf{t}_\mathsf{c})(\mathbf{r}^\mathsf{T},\ 0)^\mathsf{T} = \mathsf{K}\mathbf{r}$. The fixed point property follows as

$$\mathsf{H}_\infty\mathbf{a} = \mathsf{K}\mathsf{R}_\mathsf{c}\mathsf{K}^{-1}\mathbf{a} = \mathsf{K}\mathsf{R}_\mathsf{c}\mathbf{r} = \mathsf{K}\mathbf{r} = \mathbf{a}$$

and similarly for $\mathbf{I_a}$ and $\mathbf{J_a}$.

Also, the eigenvectors of $\mathsf{H}_\infty^\mathsf{T}$ are the vanishing line of the plane orthogonal to the rotation axis, $\mathrm{l}_\mathbf{a} = \mathbf{I_a} \times \mathbf{J_a}$, and the lines $\mathrm{l}_\mathbf{I} = \mathbf{a} \times \mathbf{I_a}$ and $\mathrm{l}_\mathbf{J} = \mathbf{a} \times \mathbf{J_a}$.

Constraints on ω from these entities (see figure 5.17) are firstly, that $\mathbf{I_a}$ and $\mathbf{J_a}$ are imaged circular points and thus lie on ω, so

$$\mathbf{I_a}^\mathsf{T}\omega\mathbf{I_a} = 0 \ \text{ and } \ \mathbf{J_a}^\mathsf{T}\omega\mathbf{J_a} = 0 \tag{5.2}$$

Secondly, $\mathbf{I_a}$ and $\mathbf{J_a}$ are both conjugate to $\mathbf{a}$ since they are the images of orthogonal ideal points:

$$\mathbf{I_a}^\mathsf{T}\omega\mathbf{a} = 0 \ \text{ and } \ \mathbf{J_a}^\mathsf{T}\omega\mathbf{a} = 0 \tag{5.3}$$

The four linear calibration constraints from the infinite homography between two views with identical cameras leaves a one parameter ambiguity in ω. The ambiguity can be parametrized by the conics described by $\mathbf{l_a}$,$\mathbf{l_I}$ and $\mathbf{l_J}$ [117]

$$\mathsf{C}(\rho,\tau) = \rho\, \mathbf{l_a}\mathbf{l_a}^\mathsf{T} + \tau\, (\mathbf{l_I}\mathbf{l_J}^\mathsf{T} + \mathbf{l_J}\mathbf{l_I}^\mathsf{T}) = \rho\, \mathsf{L}_\mathbf{a} + \tau\, \mathsf{L}_\mathbf{IJ} \tag{5.4}$$

$\mathsf{C}(\rho,\tau)$ satisfies both (5.2) and (5.3).

The one parameter ambiguity in ω can be resolved by supplying information about camera parameters such as the aspect ratio, skew or principal point. Degenerate conditions under which specifying one of these does not resolve the ambiguity have been analysed in Zisserman *et al*[117]. The following section goes on to consider the combination of the fixed camera constraints and the constraints obtained from the vanishing points of orthogonal directions.

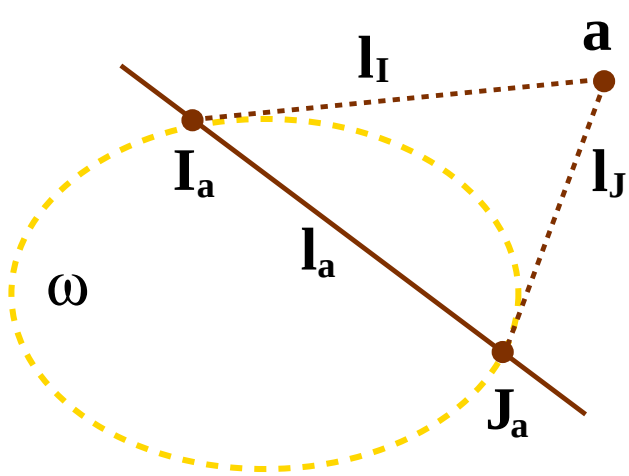


Figure 5.17: Four constraints on ω can be written from the fixed points of H_∞. The vanishing point of the axis of rotation between cameras, $\mathbf{a}$, is conjugate to both the imaged circular points of the plane orthogonal to the axis, $\mathbf{I_a}$ and $\mathbf{J_a}$. The latter also lie on ω.

5.4.2 Identical camera degeneracies

The four calibration constraints obtained from identical cameras require knowledge of H_∞, which can be computed from the fundamental matrix and three matched vanishing points in two views.

When the vanishing points represent orthogonal directions, conjugacy constraints on ω are also available in both views. The combination of the conjugacy constraints and fixed camera constraints in general provide up to seven constraints on ω, four from (5.2) and (5.3), and three from (4.2) for the three vanishing point pairs. This section examines degenerate cases by applying the conjugacy constraints to the one parameter family $\mathsf{C}(\rho, \tau)$ from (5.4). This is in preference to considering the two parameter family of conics defined by the triad of vanishing points, and investigating the degeneracy associated with the four constraints on two degrees of freedom. In the former case, it is necessary to consider three constraints on one degree of freedom. For vanishing points $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$, the constraints are

$$\begin{aligned} \mathbf{v}_1^\top(\rho\, \mathbf{l_a}\mathbf{l_a}^\top + \tau\, (\mathbf{l_I}\mathbf{l_J}^\top + \mathbf{l_J}\mathbf{l_I}^\top))\mathbf{v}_2 &= 0 \\ \mathbf{v}_1^\top(\rho\, \mathbf{l_a}\mathbf{l_a}^\top + \tau\, (\mathbf{l_I}\mathbf{l_J}^\top + \mathbf{l_J}\mathbf{l_I}^\top))\mathbf{v}_3 &= 0 \\ \mathbf{v}_2^\top(\rho\, \mathbf{l_a}\mathbf{l_a}^\top + \tau\, (\mathbf{l_I}\mathbf{l_J}^\top + \mathbf{l_J}\mathbf{l_I}^\top))\mathbf{v}_3 &= 0 \end{aligned} \tag{5.5}$$

Two degenerate cases are enumerated. In the first it is shown that two of the three constraints in (5.5) are satisfied for all ρ and τ, and thus do not constrain these parameters. One constraint remains.

In the second case, all three constraints in (5.5) are satisfied for all ρ and τ, and no constraints are obtained. Writing (5.5) in matrix form with coefficients of ρ and τ in matrix A

$$\mathsf{A}\begin{pmatrix}\rho \\ \tau\end{pmatrix} = \mathbf{0}$$

A has rank one in the first case and rank zero in the second.

The first degenerate condition occurs when $\mathbf{l_a}$ intersects one vertex of the self-polar triangle, say $\mathbf{v}_1$. The first term in each of the first two constraints in (5.5), the two involving $\mathbf{v}_1$, is zero, since $\mathbf{v}_1^\top \mathbf{l_a} = 0$. Also, the polar of $\mathbf{v}_1$ with respect to ω is

$$\mathbf{l}_1 = \omega \mathbf{v}_1 = \mathsf{L}_{\mathbf{IJ}}\mathbf{v}_1 = \mathbf{v}_2 \times \mathbf{v}_3$$

so $\mathbf{v}_1$ is conjugate to both $\mathbf{v}_2$ and $\mathbf{v}_3$ with respect to $\mathsf{L}_{\mathbf{IJ}}$, and

$$\begin{aligned}\mathbf{v}_1^\top(\mathbf{l}_\mathbf{I}\mathbf{l}_\mathbf{J}^\top + \mathbf{l}_\mathbf{J}\mathbf{l}_\mathbf{I}^\top)\mathbf{v}_2 &= 0\\ \mathbf{v}_1^\top(\mathbf{l}_\mathbf{I}\mathbf{l}_\mathbf{J}^\top + \mathbf{l}_\mathbf{J}\mathbf{l}_\mathbf{I}^\top)\mathbf{v}_3 &= 0\end{aligned}$$

leaving only the third conjugacy constraint from (5.5):

$$\mathbf{v}_2^\top(\rho\, \mathbf{l}_\mathbf{a}\mathbf{l}_\mathbf{a}^\top + \tau\, (\mathbf{l}_\mathbf{I}\mathbf{l}_\mathbf{J}^\top + \mathbf{l}_\mathbf{J}\mathbf{l}_\mathbf{I}^\top))\mathbf{v}_3 = 0 \tag{5.6}$$

There are thus a total of five constraints on ω in this case.

The second degenerate case occurs when more than one vertex of the self-polar triangle lies on $\mathbf{l}_\mathbf{a}$. Two vertices on $\mathbf{l}_\mathbf{a}$ implies that the third vertex coincides with $\mathbf{a}$: if $\mathbf{a}$ coincides with, say $\mathbf{v}_3$, then, since $\mathbf{v}_1$ and $\mathbf{v}_2$ are conjugate to $\mathbf{v}_3$ with respect to ω, they are conjugate to $\mathbf{a}$ and lie on $\mathbf{l}_\mathbf{a}$. Hence $\mathbf{v}_2^\top\mathbf{l}_\mathbf{a} = 0$ and $\mathbf{v}_3^\top\mathbf{l}_\mathbf{I} = \mathbf{v}_3^\top\mathbf{l}_\mathbf{J} = \mathbf{a}^\top\mathbf{l}_\mathbf{I} = 0$. Consequently,

$$\mathbf{v}_2^\top(\rho\, \mathbf{l}_\mathbf{a}\mathbf{l}_\mathbf{a}^\top + \tau\, (\mathbf{l}_\mathbf{I}\mathbf{l}_\mathbf{J}^\top + \mathbf{l}_\mathbf{J}\mathbf{l}_\mathbf{I}^\top))\mathbf{v}_3 = 0\rho + 0\tau$$

The triad of vanishing points thus places no constraints on ω additional to those obtained from the infinite homography, for a total of four constraints.

The physical configurations leading to degeneracy depend on the orientations of the cameras relative to the three orthogonal space directions. The first case, where there are five constraints on ω and calibration is possible, occurs when the plane orthogonal to the axis of the rotation between cameras is part of the pencil of planes parallel to one of the orthogonal triad. Its vanishing line then intersects a vanishing point. If the axis of rotation between cameras is the same as one of the three orthogonal directions, the second case is encountered, and only four constraints in total on ω are available. The example of figure 5.1 is such a case, with a vertical axis of rotation between views, and parallel lines in the vertical direction part of the triad. Camera specific information, such as the aspect ratio, skew and principal point could then be used to constrain ω.

5.5 Summary

This chapter has presented an extension of the calibration method using vanishing points to two views. The approach is based on the stratified rectification of a projective reconstruction obtained from a fundamental matrix. Three matched vanishing points and the fundamental matrix define the infinite homography and the plane at infinity, and thus an affine reconstruction. Subsequent metric rectification follows from computation of the internal parameters of the cameras via the image of the absolute conic. Constraints on the IAC are obtained from conjugacy constraints from the vanishing points of orthogonal directions. Additional constraints arise from the motion between views: the transfer of the image plane circular points between views for square pixel cameras and the fixed points of the infinite homography for identical cameras. Degenerate cases for both of these configurations have been presented.

A table summarising the degenerate conditions associated with a triad of vanishing points for orthogonal direction analysed thus far appears below. It includes single view square pixel cameras, two view square pixel cameras and two view fixed cameras.

Vanishing points	Cameras	Independent constraints
3 finite	1 square pixel	5
1 ideal	1 square pixel	4
2 ideal	1 square pixel	4
3 finite $\mathbf{l}_\infty^{*\mathsf{T}}\mathbf{v}_i \neq 0$	2 square pixel	7
3 finite $\mathbf{l}_\infty^{*\mathsf{T}}\mathbf{v}_i = 0$	2 square pixel	6
1 or 2 ideal $\mathbf{l}_\infty^{*\mathsf{T}}\mathbf{v}_i \neq 0$	2 square pixel	6
1 or 2 ideal $\mathbf{l}_\infty^{*\mathsf{T}}\mathbf{v}_i = 0$	2 square pixel	5
$\mathbf{l}_\mathbf{a}^\mathsf{T}\mathbf{v}_i \neq 0$	2 identical	7
$\mathbf{l}_\mathbf{a}^\mathsf{T}\mathbf{v}_i = 0$	2 identical	5
$\mathbf{l}_\mathbf{a}^\mathsf{T}\mathbf{v}_i = 0$ and $\mathbf{l}_\mathbf{a} = \mathbf{l}_i$	2 identical	4

Table 5.1: Constraints on ω for various combinations of cameras and motion.

Chapter 6

Uncalibrated Motion Capture Using Affine Cameras

6.1 Introduction

This chapter presents a constraint on the rectification of an affine reconstruction derived from a known ratio of world lengths. The constraint is applied to the rigid links of an articulated structure – the fixed lengths of human body components – through a motion sequence, leading to an uncalibrated motion capture system[65].

The human skeleton is commonly modeled as a rigid link articulated structure. A body segment, such as the upper arm, is treated as a link of fixed length between the shoulder and elbow joints. As a person moves, the relative orientation of all the body components changes, but the lengths of the parts remain fixed (see figure 6.1). It will be shown that the constant length of each segment of the body provides a linear constraint on the absolute conic in an affine reconstruction of a set of postures.

This material is a significant departure from the dominant theme of previous chapters. The emphasis thus far has been on calibration and reconstruction from constraints of parallelism and orthogonality, the predominant architectural features of built environments. The rigid link motion capture system, however, is used in the context of dynamic human activity.

(a) (b)

Figure 6.1: Body segments, such as the upper arm between the shoulder and elbow, and the upper leg, between the hip and knee, have a constant length when a person moves.

Motion capture itself, of course, is nothing new. There are numerous commercial systems that accurately capture the behaviour of a moving human subject using optical, magnetic and mechanical devices. These systems have a broad range of applications, including graphics animation in the film and game industries, athletic performance evaluation and clinical gait analysis. Most optical systems (such as those developed by Vicon and Peak Performance) employ large sets of cameras in a calibrated environment, and use reflective markers placed on significant joints. The (non-trivial) motion capture problem is then one of accurately tracking the markers, allowing for occlusion, and computing a calibrated reconstruction. Video based systems, (used, for example, by Ariel and Peak Performance) require spatial calibration from the scene or camera sensors. The system presented here differs in that the cameras are totally uncalibrated. The main application for the system is in broadcast sports events. In sports played on a large field there are typically a small number of cameras able to view the action in a specific part of the stadium. Cameras are often also carried by hand, and thus cannot be statically calibrated. Furthermore, cameras at the edge of a sports field, or mounted in a stadium, are usually sufficiently far from the subject of interest for an *affine* camera model to be assumed.

Research in tracking and reconstructing human motion has taken a number of directions, a few examples of which are described below. For a comprehensive review, see Gavrila[38].

Leventon and Freeman[64] take a statistical approach to motion capture from monocular se-

quences by using a motion capture database to build a probability model for motion sequences. Pose of the body in a novel sequence is then inferred by finding the most probable 3D configuration given the imaged co-ordinates. Taylor[102] assumes known size body segments imaged by a scaled orthographic camera, and computes 3D orientation of the segments from foreshortening in a single view. Webb and Aggarwal[109] consider the rotation axis of relative joints to be fixed for short times and recover 3D structure from monocular sequences.

Bregler and Malik[9] track and reconstruct human motion from single or multiple sequences using a kinematic chain model. Joint pose is parametrized by a rotation around and translation along a screw axis, and is combined with an exponential representation of rotation to describe the kinematic chains. Deutscher *et al*[26] apply particle filtering to tracking and motion recovery, while Sinclair *et al*[95] derive constraints for points on an axis of articulation in multiple views.

In this chapter the tracking problem is ignored: the objective is a calibration method which exploits the structure of the body as a source of constraints. The constraint is derived in section 6.2 in a general form based on a known ratio of 3D lengths. A number of degenerate configurations are described. The method is demonstrated in subsequent sections on sequences captured by a pair of affine cameras with manual selection of data in the images. Accordingly, section 6.3 describes the affine camera model and affine reconstruction from uncalibrated cameras. Section 6.4 discusses auto-calibration of affine cameras and explores some of the differences between pinhole and affine cameras. Finally, section 6.5 presents the motion capture method and show results obtained in reconstructing the motion of athletes from broadcast video.

6.2 Rectification constraints from relative length

This section describes a constraint on the metric rectification of an affine reconstruction derived from a known ratio of world 3D lengths. In (2.11), the rectification affinity relates point $\mathbf{X}_\mathsf{m}$ in the metric reconstruction and $\mathbf{X}_\mathsf{a}$ in the affine reconstruction:

$$\mathbf{X}_\mathsf{m} = \begin{pmatrix} \mathsf{U} & \mathbf{0}^\mathsf{T} \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix} \mathbf{X}_\mathsf{a} \tag{6.1}$$

with

$$\Omega_\infty = \mathsf{U}^\mathsf{T}\mathsf{U}$$

Assuming X_a is finite and normalizing to $\mathbf{X}_\mathsf{a} = (\tilde{\mathbf{X}}_\mathsf{a}^\mathsf{T},\ 1)^\mathsf{T}$, (6.1) reduces to

$$\mathbf{X}_\mathsf{m} = \begin{pmatrix} \mathsf{U}\tilde{\mathbf{X}}_\mathsf{a} \\ 1 \end{pmatrix}$$

Given two points $\mathbf{X}_{\mathbf{A}_1}$ and $\mathbf{X}_{\mathbf{A}_2}$ which are the endpoints of a line segment $\mathcal{L}_\mathbf{A}$ in an affine reconstruction, consider the difference vector

$$\begin{pmatrix} \mathsf{U}\tilde{\mathbf{X}}_{\mathbf{A}_1} \\ 1 \end{pmatrix} - \begin{pmatrix} \mathsf{U}\tilde{\mathbf{X}}_{\mathbf{A}_2} \\ 1 \end{pmatrix} = \begin{pmatrix} \mathsf{U}(\tilde{\mathbf{X}}_{\mathbf{A}_1} - \tilde{\mathbf{X}}\mathbf{A}_2) \\ 0 \end{pmatrix} = \begin{pmatrix} \mathsf{U}\mathbf{D}_\mathbf{A} \\ 0 \end{pmatrix}$$

$\mathbf{D}_\mathbf{A}$ is a scaled vector in the direction of the line through $\mathbf{X}_{\mathbf{A}_1}$ and $\mathbf{X}_{\mathbf{A}_2}$. The magnitude of this vector is the *metric* length of the line segment $\mathcal{L}_\mathbf{A}$.

The relative length constraint is applied when the ratio of lengths of a *pair* of line segments are known in a metric reconstruction. It will now be shown how this leads to a linear constraint on the elements of the absolute conic Ω_∞. In the methods in chapter 5 the affine transformation that metric rectifies a reconstruction is obtained by computing the *image* of the absolute conic. Here, however, it is necessary to compute the absolute conic itself. This is because the constraint is explicitly 3 dimensional and cannot be applied in an image to constrain ω.

6.2.1 The relative length constraint

Assume two 3D line segments in an affine reconstruction, segment $\mathcal{L}_\mathbf{A}$ with endpoints $\mathbf{X}_{\mathbf{A}_1}$ and $\mathbf{X}_{\mathbf{A}_2}$, and segment $\mathcal{L}_\mathbf{B}$ with endpoints $\mathbf{X}_{\mathbf{B}_1}$ and $\mathbf{X}_{\mathbf{B}_2}$. The metric lengths of $\mathcal{L}_\mathbf{A}$ and $\mathcal{L}_\mathbf{B}$, $l_\mathbf{A}$ and

$l_{\mathbf{B}}$, may be expressed by transforming the endpoints with (6.1) and taking inner products:

$$\begin{aligned} l_{\mathbf{A}}^2 &= (\mathsf{U}\mathbf{D_A})^\mathsf{T}(\mathsf{U}\mathbf{D_A}) \quad \text{and} \\ l_{\mathbf{B}}^2 &= (\mathsf{U}\mathbf{D_B})^\mathsf{T}(\mathsf{U}\mathbf{D_B}) \end{aligned} \tag{6.2}$$

If the metric length ratio of $\mathcal{L}_{\mathbf{A}}$ and $\mathcal{L}_{\mathbf{B}}$ is λ, from (6.2):

$$\begin{aligned} l_{\mathbf{A}}^2 &= \lambda^2 l_{\mathbf{B}}^2 \\ (\mathsf{U}\mathbf{D_A})^\mathsf{T}(\mathsf{U}\mathbf{D_A}) &= \lambda^2(\mathsf{U}\mathbf{D_B})^\mathsf{T}(\mathsf{U}\mathbf{D_B}) \\ \mathbf{D}_{\mathbf{A}}^\mathsf{T}\Omega_\infty\mathbf{D_A} &= \lambda^2\mathbf{D}_{\mathbf{B}}^\mathsf{T}\Omega_\infty\mathbf{D_B} \end{aligned} \tag{6.3}$$

where Ω_∞ is the symmetric 3×3 matrix with elements

$$\Omega_\infty = \begin{pmatrix} \Omega_1 & \Omega_2 & \Omega_4 \\ \Omega_2 & \Omega_3 & \Omega_5 \\ \Omega_4 & \Omega_5 & \Omega_6 \end{pmatrix} \tag{6.4}$$

Expanding (6.3) in terms of the elements of Ω_∞, $\mathbf{D_A} = (D_{\mathbf{A}1},\ D_{\mathbf{A}2},\ D_{\mathbf{A}3})^\mathsf{T}$ and $\mathbf{D_B} = (D_{\mathbf{B}1},\ D_{\mathbf{B}2},\ D_{\mathbf{B}3})^\mathsf{T}$

$$\begin{aligned} \mathbf{D}_{\mathbf{A}}^\mathsf{T}\Omega_\infty\mathbf{D_A} - \lambda^2\mathbf{D}_{\mathbf{B}}^\mathsf{T}\Omega_\infty\mathbf{D_B} &= 0 \\ (D_{\mathbf{A}1}^2 - \lambda^2 D_{\mathbf{B}1}^2)\Omega_1 + 2(D_{\mathbf{A}1}D_{\mathbf{A}2} - \lambda^2 D_{\mathbf{B}1}D_{\mathbf{B}2})\Omega_2 & \\ +(D_{\mathbf{A}2}^2 - \lambda^2 D_{\mathbf{B}2}^2)\Omega_3 + 2(D_{\mathbf{A}1}D_{\mathbf{A}3} - \lambda^2 D_{\mathbf{B}1}D_{\mathbf{B}3})\Omega_4 & \\ +2(D_{\mathbf{A}2}D_{\mathbf{A}3} - \lambda^2 D_{\mathbf{B}2}D_{\mathbf{B}3})\Omega_5 + (D_{\mathbf{A}3}^2 - \lambda^2 D_{\mathbf{B}3}^2)\Omega_6 &= 0 \end{aligned}$$

This is a linear equation in the elements of Ω_∞, which we may write in vector form as

$$\mathbf{c}^\mathsf{T}\mathbf{\Omega}_\mathsf{v} = 0 \tag{6.5}$$

where $\mathbf{c} = (c_1,\ \ldots,\ c_6)^\mathsf{T}$ is the vector of coefficients and $\mathbf{\Omega}_\mathsf{v} = (\Omega_1,\ \ldots,\ \Omega_6)^\mathsf{T}$ is the vector of elements of Ω_∞.

As before with ω, Ω_∞ is a homogeneous matrix with five degrees of freedom. Five independent constraints are thus required to solve for Ω_∞. Given five or more independent constraints, the coefficient vectors $\mathbf{c}_1$ to $\mathbf{c}_n$ can be combined in a constraint matrix. It follows from (6.5) that

$$\begin{pmatrix} \mathbf{c}_1^\mathsf{T} \\ \vdots \\ \mathbf{c}_n^\mathsf{T} \end{pmatrix} \boldsymbol{\Omega}_\mathsf{v} = \mathsf{C}\boldsymbol{\Omega}_\mathsf{v} = \mathbf{0} \tag{6.6}$$

and $\boldsymbol{\Omega}_\mathsf{v}$ is the null vector of C, which is of rank five.

In the presence of noise and with more than five constraints, C will generally be of rank six. In this case, singular value decomposition (SVD) provides an estimate of the null vector. The singular vector associated with the smallest singular value is the estimate of the null vector of C that minimizes $\|\mathsf{C}\boldsymbol{\Omega}_\mathsf{v}\|$.

There are also degenerate configurations of lines segments which do not yield independent constraints. These are examined in the following section.

The relative length constraint has a (somewhat untidy) geometric interpretation. The term $\tilde{\mathbf{X}}^\mathsf{T}\Omega_\infty\tilde{\mathbf{X}}$ for a point $\tilde{\mathbf{X}}$ on $\boldsymbol{\pi}_\infty$ is the distance of $\tilde{\mathbf{X}}$ from its polar with respect to Ω_∞ (the line $\Omega_\infty\tilde{\mathbf{X}}$), scaled by a factor dependent on the magnitude of $\tilde{\mathbf{X}}$. From (6.3), the ratio of these scaled distances is an affine invariant.

6.2.2 Degeneracies

Constraint degeneracy occurs when a particular constraint fails to provide information about Ω_∞, or when a set of constraints are linearly dependent in (6.5). Three such conditions are enumerated below. They are similar to the degenerate conditions on rectifying an affine plane in chapter 3.

Parallel line segments. Length ratio in parallel directions is an affine invariant, so parallel line segments in an affine reconstruction preserve their metric length ratio. It is therefore unsurprising that no constraint on Ω_∞ is obtained from the relative length of parallel segments. Explicitly, if $\mathcal{L}_\mathbf{A}$

and $\mathcal{L}_{\mathbf{B}}$ are parallel with length ratio λ, $\mathbf{D_A} = \lambda \mathbf{D_B}$ (see figure 6.2). and (6.3) becomes

$$\begin{aligned}&\mathbf{D_A^\top}\Omega_\infty \mathbf{D_A} - \lambda^2 \mathbf{D_B^\top}\Omega_\infty \mathbf{D_B}\\ &= \lambda^2 \mathbf{D_B^\top}\Omega_\infty \mathbf{D_B} - \lambda^2 \mathbf{D_B^\top}\Omega_\infty \mathbf{D_B} = 0\end{aligned}$$

and is satisfied for all Ω_∞. The elements of the coefficient vector $\mathbf{c}$ will all be zero.

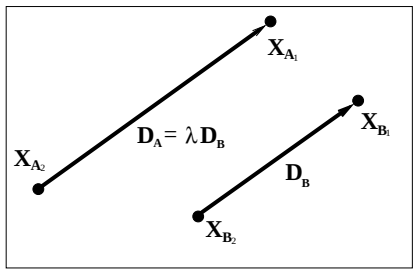


Figure 6.2: Parallel line segments provide no constraint on Ω_∞.

Pairs of parallel line segments. Two constraints obtained from the relative lengths of segment *pairs* that are parallel are not independent. Suppose we have $\mathcal{L}_{\mathbf{A}}$ and $\mathcal{L}_{\mathbf{B}}$ with metric length ratio λ, and $\mathcal{L}_{\mathbf{A}'}$ and $\mathcal{L}_{\mathbf{B}'}$ with ratio κ, shown in figure 6.3, we can write two linear constraints on Ω_∞:

$$\begin{aligned}\mathbf{D_A^\top}\Omega_\infty \mathbf{D_A} - \lambda^2 \mathbf{D_B^\top}\Omega_\infty \mathbf{D_B} &= 0\\ \mathbf{D_{A'}^\top}\Omega_\infty \mathbf{D_{A'}} - \kappa^2 \mathbf{D_{B'}^\top}\Omega_\infty \mathbf{D_{B'}} &= 0\end{aligned}$$

However, if $\mathcal{L}_{\mathbf{A}}$ is parallel to $\mathcal{L}_{\mathbf{A}'}$ and $\mathcal{L}_{\mathbf{B}}$ is parallel to $\mathcal{L}_{\mathbf{B}'}$, $\mathbf{D_{A'}} = \alpha \mathbf{D_A}$ and $\mathbf{D_{B'}} = \beta \mathbf{D_B}$. Additionally, for parallel directions length ratios are preserved, so $\frac{\alpha}{\beta} = \frac{\kappa}{\lambda}$. The pair of constraints becomes

$$\begin{aligned}\mathbf{D_A^\top}\Omega_\infty \mathbf{D_A} - \lambda^2 \mathbf{D_B^\top}\Omega_\infty \mathbf{D_B} &= 0\\ \alpha^2 \mathbf{D_A^\top}\Omega_\infty \mathbf{D_A} - \alpha^2\lambda^2\, \mathbf{D_B^\top}\Omega_\infty \mathbf{D_B} &= 0\end{aligned}$$

These two equations are linearly dependent and only one constraint on Ω_∞ is obtained.

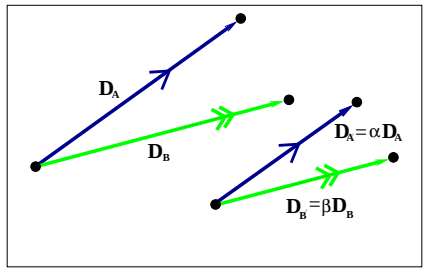


Figure 6.3: Parallel pairs of line segments provide one constraint on Ω_∞.

Co-planar or parallel plane line segments. Any number of constraints originating with coplanar line segments or line segments in parallel planes, as in figure 6.4, provide only two independent constraints on Ω_∞. Geometrically, the constraints fully define metric structure on a plane (as in chapter 3), and indeed on the pencil of parallel planes, but provide no information about other directions.

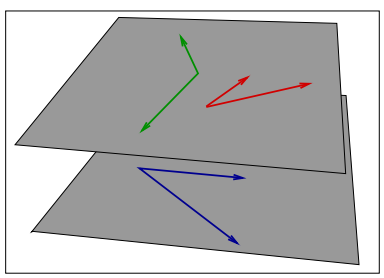


Figure 6.4: Coplanar or parallel plane line segments provided at most two constraints on Ω_∞.

As always with degenerate conditions, care must be taken in implementing the constraints. Theoretically degenerate constraints may appear to provide valid constraints due to measurement noise, and near degenerate conditions provide unstable results.

6.2.3 The rigid link constraint

The relative length constraint can be applied in practice without quantitative knowledge of length ratios. This is the case when the qualitative observation can be made that objects of equal length are

present in a scene, determining a unit length ratio constraint. This is the idea in applying it to human motion as the movement of an articulated rigid link structure. The sequence of postures sampled by a pair of cameras viewing the motion allows reconstruction of the postures. Treated as a static 3D object, the set of postures then contains a number of fixed length objects: each body segment, such as the upper arm, remains a constant length as it moves. This is the rigid link constraint.

Given an affine reconstruction of a human motion, constraints on metric rectification can be written from the fixed length of a number of body segments. Later sections detail the use of these constraints, but first it is necessary to examine the affine camera model since the application which follows assumes the use of affine cameras.

6.3 Affine cameras and reconstruction

6.3.1 The affine camera model

Affine cameras have a number of properties that distinguish them from their perspective counterparts. General affine cameras are characterized by an optic centre at infinity[73]. Using the affine camera model, world points are imaged by *parallel* projection (figure 6.5). The rays through the camera centre, the world points and the image points are a parallel family. The imaged co-ordinates of a world point are thus independent of its depth in the scene.

The affine projection matrix may be written in a form similar to the general camera projection matrix. The canonical affine camera, aligned with the world co-ordinate system, is described by

$$\mathsf{P}_\mathsf{A} = \begin{pmatrix} \mathsf{K}_\mathsf{A} & \mathbf{0} \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} s & k & 0 \\ 0 & rs & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

where K_A is the camera internal parameter matrix. The parameter s is an isotropic scaling of image co-ordinates. The isotropic scaling can account for the changes in the imaged size of an object as a physical camera zooms or translates towards or away from it. Skew k and aspect ratio r model the skew and stretch of the optic array as before. Note that the principal point is not defined for an

affine camera. Various form of K_A define a hierarchy of cameras with an infinite optic centre[49]:

1. orthographic projection, with K_A the identity matrix,
2. scaled orthographic projection for $k = 0$ and $r = 1$, analogous to the square pixel projective camera,
3. weak perspective projection, with $k = 0$ and $r \neq 1$, and
4. the general affine camera, defined with no restrictions on the parameters of K_A.

The role of internal parameters in rectification of affine camera reconstructions will be addressed in section 6.4.

A camera with pose described by translation vector $\mathbf{t} = (t_1,\ t_2,\ t_3)^\mathsf{T}$ and rotation matrix R with rows $\mathbf{r}_1^\mathsf{T}$, $\mathbf{r}_2^\mathsf{T}$ and $\mathbf{r}_3^\mathsf{T}$, is written

$$\mathsf{P}_\mathsf{A} = \begin{pmatrix} \mathsf{K}_\mathsf{A} & \mathbf{0} \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix} \begin{pmatrix} \mathbf{r}_1^\mathsf{T} & t_1 \\ \mathbf{r}_2^\mathsf{T} & t_2 \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix}$$

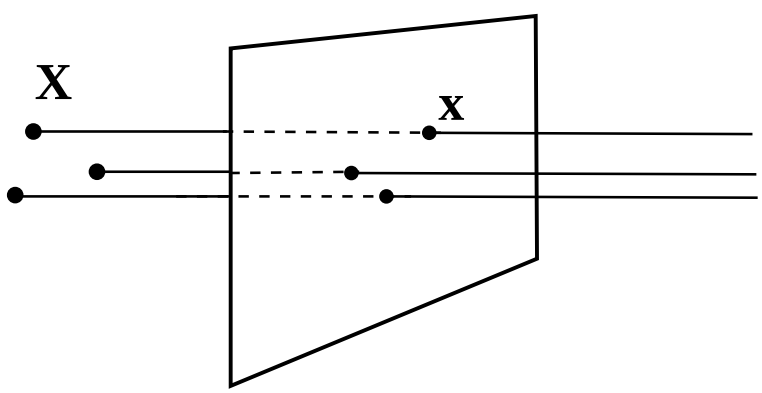


Figure 6.5: The centre of projection of an affine camera is at infinity. World points are projected orthographically onto the image plane.

Finally, note that the third row of zeros of the affine projection matrix may be neglected in an

inhomogeneous representation:

$$\tilde{\mathbf{x}} = \mathsf{K}_\mathsf{A} \begin{pmatrix} \mathbf{r}_1^\top \\ \mathbf{r}_2^\top \end{pmatrix} \tilde{\mathbf{X}} + \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} = \mathsf{G}\tilde{\mathbf{X}} + \tilde{\mathbf{t}} \tag{6.7}$$

6.3.2 Affine reconstruction

Given two affine views of scene, world geometry can be reconstructed up to an affine transformation. This result, originating with Koenderink and van Doorn[62], requires four points matched across views to define an affine basis for reconstruction. Further points may then be given 3D co-ordinates with respect to these basis vectors. A minimum of four points is also required to compute the affine fundamental matrix[115] and thus affine epipolar geometry.

The most widely used affine reconstruction method is the factorization algorithm of Tomasi and Kanade[103]. Factorization provides a robust and computationally efficient method of computing affine structure from point correspondences using singular value decomposition. Reid and Murray[85] have also shown that it gives Maximum Likelihood reconstruction under Gaussian noise assumptions for image points.

Using factorization, affine 3D structure is computed by the SVD of a measurement matrix W which, for two views, has the form

$$\mathsf{W} = \begin{pmatrix} \tilde{\mathbf{x}}_1 & \tilde{\mathbf{x}}_2 & \dots & \tilde{\mathbf{x}}_n \\ \tilde{\mathbf{x}}'_1 & \tilde{\mathbf{x}}'_2 & \dots & \tilde{\mathbf{x}}'_n \end{pmatrix}$$

for n inhomogeneous matched points $\tilde{\mathbf{x}}_i$ in the first view and $\tilde{\mathbf{x}}'_i$ in the second. The singular decomposition of W returns three matrices from which a pair of cameras G and G' as well as 3D points $\mathbf{X}_{\mathsf{a}i}$ can be extracted. Note that image co-ordinates are normalized in W to place the origin at the centroid of the points in each image. Since in parallel projection the 3D centroid of the point cloud is imaged at the 2D centroid, this centres the reconstruction coordinate system at the 3D centroid. In consequence, the translation parameters of the cameras are eliminated in (6.7), and G and G' are 2×3 inhomogeneous projection matrices.

The reconstruction is, of course, related to world structure by an unknown affine transformations. Rectifying the structure and cameras to metric requires calibration. Calibration of affine camera systems, however, differs markedly from that of projective cameras. The plane at infinity is already at infinity in the affine reconstruction, and the rectifying affine transformation is not the camera internal parameter matrix. Also, as the following section shows, the absolute conic is not imaged as a proper conic. The calibration process must thus consider the absolute conic itself, and not its image.

6.4 Calibration of affine cameras

Methods of constraining the internal parameters of a perspective camera described in previous chapters do not extend to affine cameras. This section shows that the plane at infinity and absolute conic behave rather differently under affine projection. Both entities project to lower dimensions, and cannot be used in calibration in the same way as their projective counterparts. This has a number of consequences for combining scene and camera information in the rectification process.

6.4.1 The geometry of parallel projection

The absolute disk quadric projects to the conic dual to the image plane circular points:

$$
\begin{aligned}
\mathsf{P}_\mathsf{A}\mathsf{Q}_\infty\mathsf{P}_\mathsf{A}^\mathsf{T} &= \begin{pmatrix} \mathsf{K}_\mathsf{A} & \mathbf{0} \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix} \begin{pmatrix} \mathbf{r}_1^\mathsf{T} & t_1 \\ \mathbf{r}_2^\mathsf{T} & t_2 \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{0} \\ t_1 & t_2 & 1 \end{pmatrix} \begin{pmatrix} \mathsf{K}_\mathsf{A}^\mathsf{T} & \mathbf{0} \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix} \\
&= \begin{pmatrix} \mathsf{K}_\mathsf{A} & \mathbf{0} \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix} \begin{pmatrix} \mathbf{r}_1^\mathsf{T}\mathbf{r}_1 & \mathbf{r}_1^\mathsf{T}\mathbf{r}_2 & 0 \\ \mathbf{r}_2^\mathsf{T}\mathbf{r}_1 & \mathbf{r}_2^\mathsf{T}\mathbf{r}_2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \mathsf{K}_\mathsf{A}^\mathsf{T} & \mathbf{0} \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix} = \begin{pmatrix} \mathsf{K}_\mathsf{A}\mathsf{K}_\mathsf{A}^\mathsf{T} & \mathbf{0} \\ \mathbf{0}^\mathsf{T} & 0 \end{pmatrix} \\
&= \mathsf{D}_\infty = \mathbf{I}\mathbf{J}^\mathsf{T} + \mathbf{J}\mathbf{I}^\mathsf{T} \qquad (6.8)
\end{aligned}
$$

It follows that the entire plane at infinity itself projects to the ideal line of the image plane $\mathbf{l}_\infty$. This can be confirmed by observing that any ideal 3D point $\mathbf{X}_\infty = (\tilde{\mathbf{X}}_\infty,\ 0)^\mathsf{T}$ projects to an ideal 2D point

$$\begin{aligned}
\mathbf{x}_\infty &= \mathsf{P_A}\mathbf{X}_\infty = \begin{pmatrix} \mathsf{K_A} & \mathbf{0} \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix} \begin{pmatrix} \mathbf{r}_1^\mathsf{T} & t_1 \\ \mathbf{r}_2^\mathsf{T} & t_2 \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix} \begin{pmatrix} \tilde{\mathbf{X}}_\infty \\ 0 \end{pmatrix} = \begin{pmatrix} \mathsf{K_A} & \mathbf{0} \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix} \begin{pmatrix} \mathbf{r}_1^\mathsf{T}\tilde{\mathbf{X}}_\infty \\ \mathbf{r}_2^\mathsf{T}\tilde{\mathbf{X}}_\infty \\ 0 \end{pmatrix} \\
&= \begin{pmatrix} \mathsf{K_A} \begin{pmatrix} \mathbf{r}_1^\mathsf{T} \\ \mathbf{r}_2^\mathsf{T} \end{pmatrix} \\ \mathbf{0}^\mathsf{T} \end{pmatrix} \tilde{\mathbf{X}}_\infty = \mathsf{T}\tilde{\mathbf{X}}_\infty
\end{aligned} \tag{6.9}$$

so T is a rank 2 degenerate mapping from the plane at infinity to the ideal line of the image plane.

Points on the absolute conic are projected onto the image plane ideal line, which can be considered the associated conic locus of the image plane circular points. The point pair dual conic and the line are, of course, not strictly dual. The point is that the types of calibration constraints on ω derived from imaged plane circular points and vanishing points in the context of a projective camera clearly cannot be applied to an affine camera.

The image plane circular points are still defined by $\mathsf{K_A}$, so they are known if the skew and aspect ratio of the camera are known. A scaled orthographic (or square pixel) camera leaves the circular points in canonical position on the ideal line of the image plane. However, imaged orthogonal vanishing points, for example, provide no single view internal camera calibration information (unless they are parallel to the image plane, and thus harmonic conjugates with respect to the circular points). Also, the rectangle ambiguity (section 3.5.3) for the rectification of an affine imaged plane cannot be resolved by internal calibration unless the world plane is parallel to the image plane.

Internal parameters and auto-calibration constraints may be included in the computation of Ω_∞. This is addressed next.

6.4.2 Affine camera auto-calibration and known parameters

In the original work on factorization[103], rectification of affine structure relies on the orthographic projection camera model. For orthographic projection, the rows of the metric rectified cameras are orthonormal. This constraint applies equally to scaled orthography: the rows of each rectified camera are orthogonal and of equal length. Writing this for the inhomogeneous camera G, a rectified camera is given by

$$\mathsf{G}\mathsf{U}^{-1} = \begin{pmatrix} \mathbf{g}_1^\intercal \\ \mathbf{g}_2^\intercal \end{pmatrix} \mathbf{U}^{-1} = \begin{pmatrix} s\mathbf{r}_1^\intercal \\ s\mathbf{r}_2^\intercal \end{pmatrix}$$

Hence

$$\mathbf{g}_1^\intercal \mathsf{U}^{-1}(\mathbf{g}_2^\intercal \mathsf{U}^{-1})^\intercal = \mathbf{g}_1^\intercal \mathsf{U}^{-1}\mathsf{U}^{-\intercal}\mathbf{g}_2 = \mathbf{g}_1^\intercal \Omega_\infty^{-1}\mathbf{g}_2 = 0 \tag{6.10}$$

and

$$\mathbf{g}_1^\intercal \mathsf{U}^{-1}(\mathbf{g}_1^\intercal \mathsf{U}^{-1})^\intercal - \mathbf{g}_2^\intercal \mathsf{U}^{-1}(\mathbf{g}_2^\intercal \mathsf{U}^{-1})^\intercal = \mathbf{g}_1^\intercal \Omega_\infty^{-1}\mathbf{g}_1 - \mathbf{g}_2^\intercal \Omega_\infty^{-1}\mathbf{g}_2 = 0 \tag{6.11}$$

Given three or more cameras, Ω_∞^{-1} can be identified and metric structure obtained. Extensions of this approach to other camera models may be found [78, 92, 110], but scaled orthography is of most interest here since this model can be used to describe a square pixel camera at some distance from the object of interest.

Note that constraints (6.10) and (6.11) are conjugacy and relative length constraints. The rows of G are the normals of 3D planes through the camera centre. These planes are the two projection planes of the camera[30] defining its vertical and horizontal axes. Equation (6.10) expresses the conjugacy of the ideal lines on $\boldsymbol{\pi}_\infty$ defined by the $\mathbf{g}_1$ and $\mathbf{g}_2$ with respect to Ω_∞^{-1}. Similarly, (6.11) is the dual form of the relative length constraint (6.3) for these lines.

A method for auto-calibration of a number of affine cameras with unknown but fixed internal parameters is described by Quan[83]. The auto-calibration constraint assumes unchanging camera parameters, and is thus an affine camera version of the fixed camera auto-calibration constraint

described in section 5.4. In geometric terms, the constraint arises since the fixed camera projects to an unchanging dual circular point conic in each image. Hence

$$\mathsf{G}_i \Omega_\infty^{-1} \mathsf{G}_i^\intercal = \mathsf{G}_j \Omega_\infty^{-1} \mathsf{G}_j^\intercal \tag{6.12}$$

for all camera pairs i, j. Quadratic constraints on the elements of Ω_∞^{-1} follow from dehomogenising (6.12) by dividing each side through by one of the entries of the matrix. Each pair of cameras provides two constraints, so a minimum of four cameras are required to solve for Ω_∞^{-1}. With knowledge of some internal parameters or more views, the constraint can be linearised.

In the motion capture system which will be described next, two cases will be considered:

1. two fixed general affine cameras with different parameters observing the entire sequence, and
2. n pairs of scaled orthographic cameras with changing parameters, each imaging one frame of the motion.

In the first case, a pair of sequences captured by unchanging cameras is equivalent to two images of the motion. Each view is regarded as having viewed the entire set of sampled postures from the motion in time in one image. The postures are treated as a static scene, requiring a single reconstruction, and rectification transformation. There are generally a large number of rigid link constraints on this rectification. If the cameras are changing, however, the postures cannot be regarded as a static set of in a single pair of images, and each frame must be reconstructed and rectified separately. There are also fewer rigid link constraints on each of these transformations, so knowledge of the camera parameters is helpful. These two cases are dealt with in detail in the following sections.

6.5 Motion capture

The rigid link constraint applied to motion capture exploits the constant length of body segments as a subject moves to constrain the metric rectification of an affine reconstruction. The affine reconstruction may be obtained from an initial projective reconstruction using various methods of computing the infinite homography. Here, however, the emphasis is on broadcast footage of sports.

In sports broadcasts, interesting events are typically shown from a second camera in the replay, providing two views of the action. The cameras at sports events are also often placed far from the participants, at the edge of the field or mounted on the stands. The relief of athletes bodies is thus generally small compared to their depth in the scene and the affine camera model is applicable. Consequently, an affine reconstruction can be created directly from matched points. The following sections describe the manual matching process, reconstruction and calibration steps used to capture the 3D motion of a person from a pair of views captured with affine cameras.

The initial method described assumes a pair of fixed cameras. This is a somewhat naive assumption for sports broadcasts, since cameras often pan and zoom while following a player. Despite this, however, a fixed camera assumption is valid in some situations, and can also result in realistic animation even when violated. This is followed by a method of jointly calibrating a number of pairs of cameras using rigid links. It is shown that, using the scaled orthography camera model, the change in scale factor resulting from a zooming camera can be eliminated, although rotation between cameras remains unknown.

6.5.1 Affine reconstruction

The first step in affine reconstruction from two views is to select the corresponding image points defining the body parts of interest. A 15 point model of the human skeleton is used, specifying the dominant joints: shoulders, elbows, hips, knees and ankles, as well as the tips of the feet and the head or neck. An example, two views of a triplejump event, appears in figure 6.6, with the skeleton points shown in two images.

This footage is captured by a pair of cameras a significant distance from the athlete. One camera is viewing the jump from the front and is zooming as the athlete moves. The zoom matches the increase in apparent size as the athlete approaches the camera, with the result that she remains approximately the same size throughout the sequence. The second camera is viewing the jump from the side, with the motion approximately parallel to the image plane, and is panning as the athlete moves across its field of view to keep her centered in the image. The camera has a long focal length and the distance from the track is large enough that the camera rotation is negligible for current

Figure 6.6: The triplejump. Frames 13, 26, 30 and 51 of an 93 frame sequence, shown in side view above and front view below. Manually selected joints are shown in the first frame of each set.

purposes. The cameras are thus treated as a pair of fixed affine cameras viewing a motion as if on a treadmill. Translation of the athlete is lost, but the relative movement of the body parts can be reconstructed.

The skeleton points are selected by hand in the images, frame by frame. Two difficulties typically occur. The first is occlusion, either self-occlusion of joints as a body moves, or occlusion from other people or objects in the scene. Self-occlusion is relatively benign, since the position of the occluded joint can often be inferred (by a human operator) from posture and the other view. Occlusion by other objects can be a severe problem, depending on the extent. A number of football players clustered in a goal area, for example, can obscure a significant part of the player of interest.

The second difficulty is image dropout. Sequences are digitised from interlaced broadcast video, and are of poor quality. Rapidly moving features, like the hands, often smear in motion blur, making precise joint location difficult.

All these features are present in the triplejump sequence. The athlete's hands move very fast and blur considerably in the side view, and the left shoulder, arm and hip are self-occluded in much of that sequence. Humans are, however, good at inferring the position of self-occluded joints, so reasonable guesses can be made about the occluded co-ordinates. Motion blur remains a source of inaccuracy of the order of a few pixels, but this is unavoidable given the nature of the data. Uncertain

points in either view are flagged in the clicking process.

Having selected the corresponding point pairs, an initial affine reconstruction is computed using the factorization algorithm. Points flagged as uncertain in the selection process can be neglected from this reconstruction. The affine fundamental matrix can then be computed from the cameras and used to recompute the uncertain points by projecting them to the epipolar line defined by the fundamental matrix and the corresponding point in the other view[93]. The system thus tolerates points uncertain in one view, but clearly visible in the other.

An affine reconstruction of the triplejump sequence is depicted in figure 6.7. Wireframe models for four frames corresponding to figure 6.6 are shown, with horizontal translation introduced. Note the effect of affine distortion, particularly on the aspect ratio of the body.

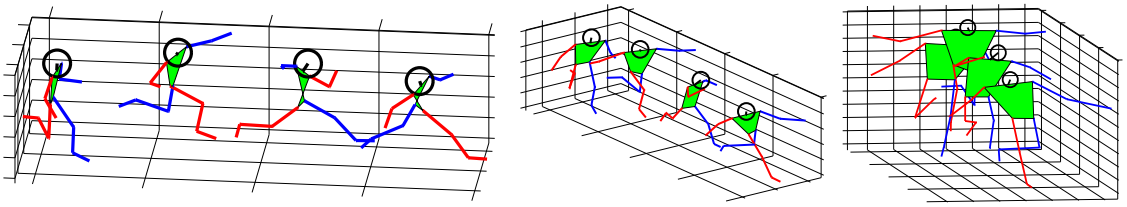


Figure 6.7: Four poses from the affine reconstruction of the triplejump sequences. The same four poses are shown from three viewpoints. Compare to the corresponding frames in figure 6.6.

The affine reconstruction can now be metric rectified using the rigid link constraint.

6.5.2 Metric rectification

The rigid link constraint is present in two forms in human motion. This first is the constant length of body segments – the upper and lower arms and legs and the distance between the hip joints – through the motion. There are nine constraints on Ω_∞ from each view pair for these particular segments. The second source of constraints is symmetry, the equal length of left and right arm and leg segments, providing four constraints per frame. Constraints are obtained from both these sources using (6.3) with a length ratio of 1, resulting in a constraint matrix of the form in (6.6). In principal, two frames are sufficient to over-constrain Ω_∞, but for interesting motions there are typically tens or hundreds of frames, so Ω_∞ is significantly over-constrained.

Note also that, in applying (6.3), the rigid link constraints between frames can be written for the

reconstruction of a single body segment between any two postures. There are, for a single segment appearing in n frames, $\binom{n}{2}$ constraints, of which $n-1$ are independent. The approach taken here is to write the $n-1$ constraints for each segment in successive pairs of frames in time.

SVD of the constraint matrix and Cholesky decomposition of the resulting estimate of Ω_∞ yield a rectification matrix U, and metric 3D structure follows from (6.1). Metric rectified structure for the triplejump example appears in figure 6.8.

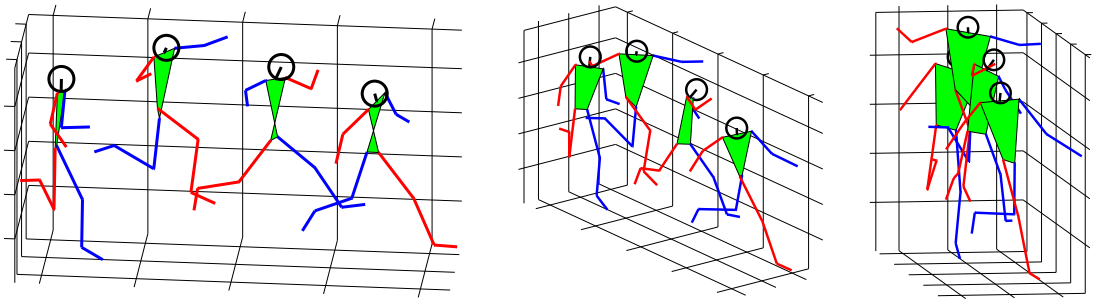


Figure 6.8: Four poses from the metric reconstruction of the triplejump sequences. The same four poses are shown from three viewpoints. Compare to the corresponding frames in figure 6.7.

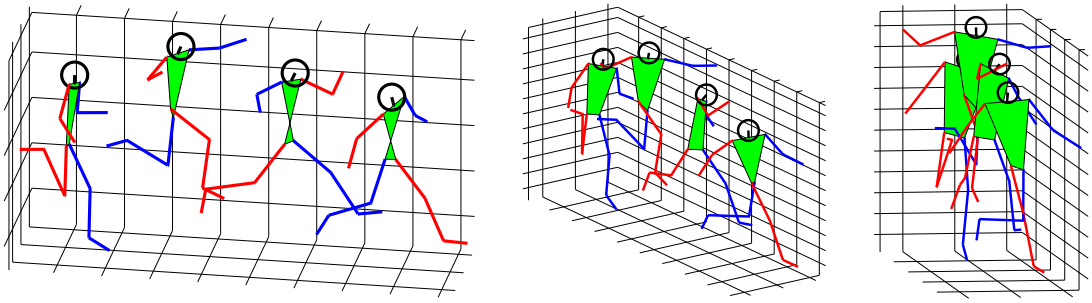


Figure 6.9: The reconstruction of the triplejump sequence, with fixed length segments fitted to body parts.

The final step in reconstructing the motion is regularization of segment sizes. Since there is noise in the system, the metric humanoid figure does not have exactly fixed length body segments throughout the reconstructed sequence. The model for each frame of the sequence is thus replaced by a fixed size humanoid with median segment lengths while preserving the angles of body segments. The results appear in figure 6.9.

Stick figures representations of motion have limited usefulness on their own. A solid skeletal model provides a far more informative representation of the motion. Ultimately, of course, the

movement can be mapped to a realistic, textured humanoid model. Accordingly, the motion capture data is exported to a format understood by a commercial animation package, the *Character Studio* plugin to *3D Studio Max*. Character Studio is able to take 3D positional motion capture data and convert it to a joint rotation hierarchy using an inverse kinematic system. The motion is mapped onto a generic biped figure, allowing further processing including manual editing, frame decimation and animation of a more complex mesh figure. The biped triplejumper appears in figure 6.10.

The motion data has also been applied to a generic VRML mesh based humanoid, shown in figure 6.11. The humanoid was created using Blaxxun Avatar Studio, a product allowing limited customization of generic humanoid meshes for use in online chat spaces with a simple set of animated gestures. The result is a low polygon mesh suitable for online use, although with limited mobility.

The quality of the reconstruction of motion is difficult to evaluate. Clearly, the accuracy of calibrated, marker based, multi-camera systems cannot be replicated. However, when viewed as an animation, the motion appears to be a faithful reproduction of the action. The animation is smooth and duplicates not only the coarse motion of the arms and legs, but also captures some of the subtle elements of the movement, such as the asymmetry in the hip and shoulder positions through the jump.

Note that degeneracy plays an important role in human motion capture. A pure translation of limbs, for example, leaves them parallel, leading to degenerate constraints. This might occur for a person walking with arms held in a fixed position holding something. Furthermore, when walking or running in a straight line, human arms and legs tend to move roughly in a pair of parallel vertical planes. The calibration constraints from rigid links in the arms and legs provide little information outside of these planes. The use of the constant distance between the hip joints is thus important (the shoulders tend to be too independently mobile to provide a similar constraint).

6.5.3 Changing cameras

The cameras in the triplejump example can be treated as fixed, with reasonable results. However, this is not always the case. The cameras filming sports events like football often zoom in on a player

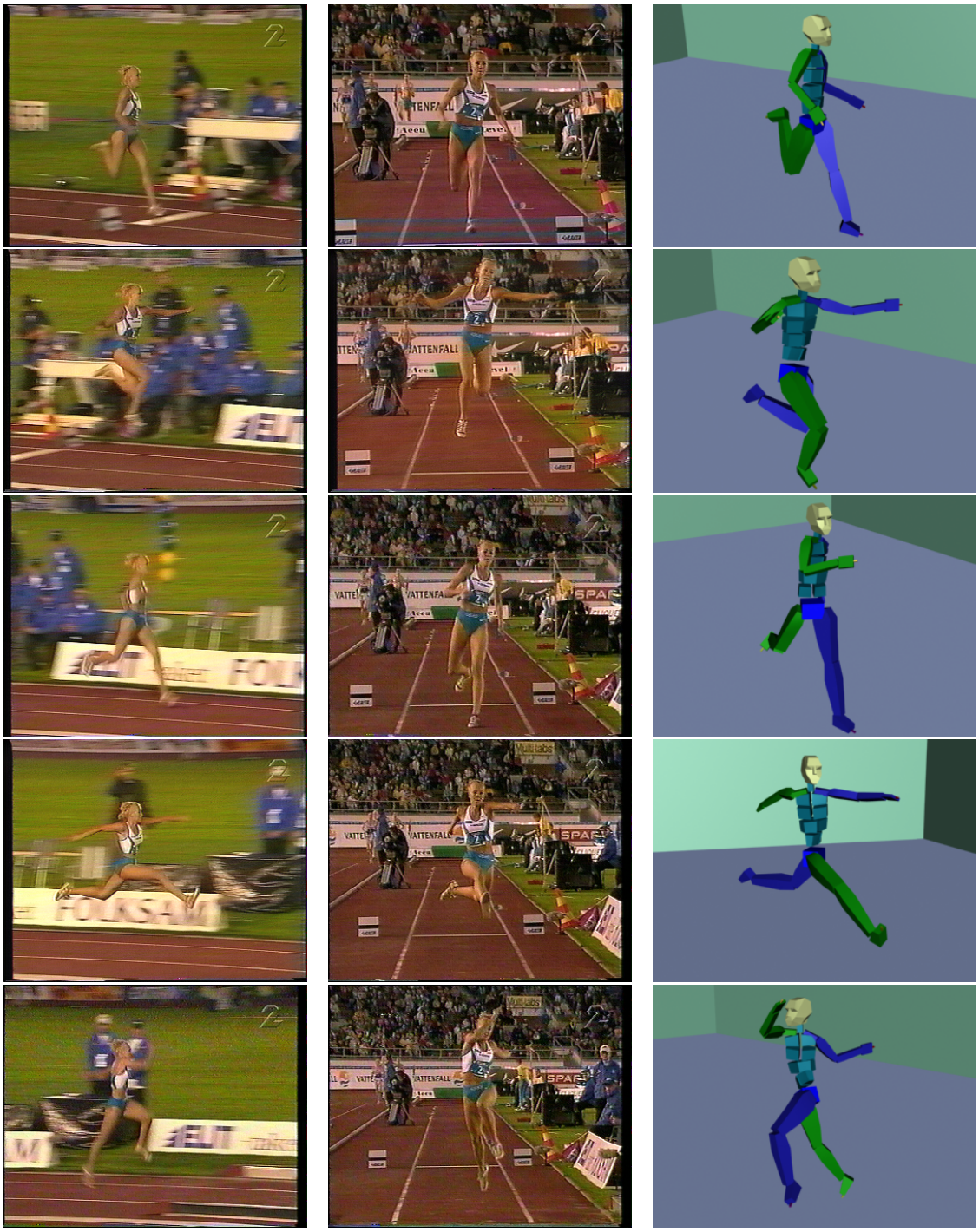


Figure 6.10: Five frames from the triplejump, equally spaced in time. In each row, frames from the two input image sequences appear alongside snapshots of an animated generic biped from various viewpoints.

Figure 6.11: Three images from the animated VRML model.

receiving the ball, and pan to follow a running player. In many situations, the rotation is necessarily limited to keep the player in the field of view, since the camera is some distance away. Also, when the player's motion does not include a large translation, the camera rotation tends to be small. As an example, consider the example of figure 6.12, which shows four frames from a sequence of a football goal. There is a small rotation and some zoom of the camera in the left view and a much larger zoom in the right.

It will be shown on this section how the zoom can be recovered by treating each frame pair as a separate reconstruction and rectifying the reconstructions simultaneously. When the rotation can be neglected, the fixed camera method can then be applied to capture the motion.

Before going on to describe the method, it is worthwhile observing the effect zooming cameras have on the results of the fixed camera algorithm. The obvious effect of zooming is to increase the imaged size of the player. The affine, metric and final reconstructions of the football example *assuming fixed cameras* appear in figures 6.13, 6.14 and 6.15. Observe that in the initial affine reconstruction there is, in addition to the affine distortion, an overall increase in body size reflecting the increasing zoom in the two sequences. Also, since there is some rotation of the cameras, a rotation of the body from posture to posture is introduced. It is interesting to note that despite this, the final motion reconstruction nevertheless appears to be quite realistic. This is because, while there is considerable zoom in the sequence, the rotation of each camera is relatively small. Zoom introduces errors in scale which are ameliorated by both the application of the rigid link constraint to successive frames rather than frames widely spaced in time, and the regularization of structure to fixed length segments. Nevertheless, it will now be shown how the effect of zooming can be eliminated from the images.

Given two sequences acquired with changing cameras, the matched points in the pair of images at each time instance should, strictly speaking, be reconstructed separately. The affine reconstructions obtained from each frame pair are related to the world by a general affine transformation, including a similarity component. There is thus an arbitrary rotation, translation and scaling in addition to the affine distortion encoded by the absolute conic relating the *metric* rectifications of each frame. It is, however, still possible to write rigid link constraints on the rectification of each reconstruction. These are constraints on the absolute conics for each reconstruction, denoted $\Omega_\infty^{(j)}$, $j = 1, \ldots, m$ for a pair of sequences of length m.

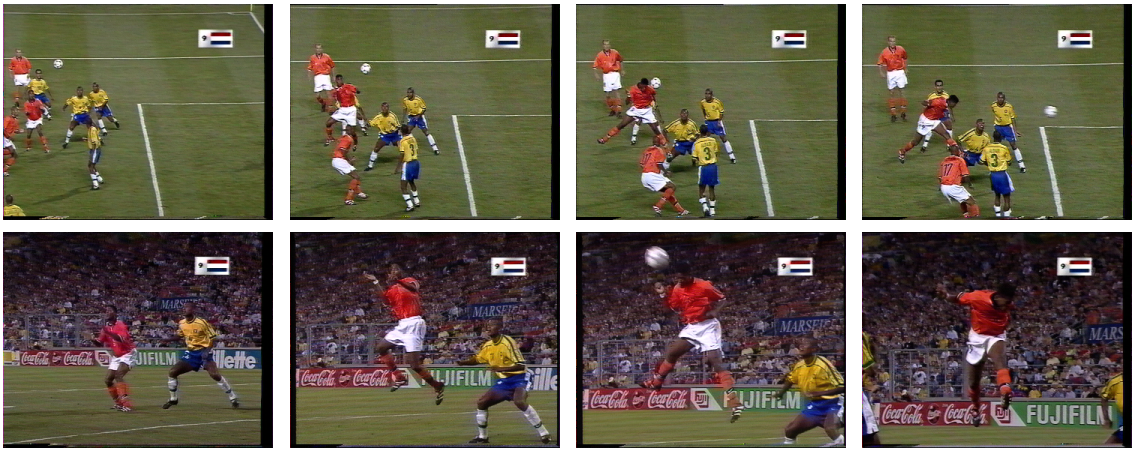


Figure 6.12: A football goal. Frames 1, 20, 30 and 40 of a 42 frame sequence, shown in two views.

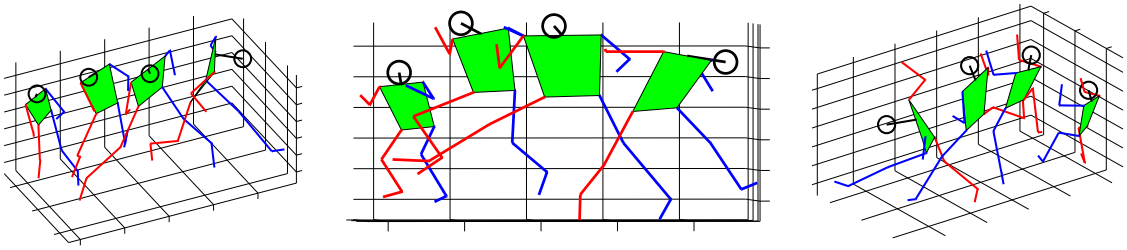


Figure 6.13: Four postures from the affine reconstruction of the sequences in figure 6.12 assuming fixed cameras.

The approach taken is to jointly calibrate all the pairs of different cameras up to a common global scale factor. Regarding the cameras as scaled orthographies, the camera scale factors are then known. The image points can be scaled accordingly, compensating for the changing zoom. The effects of panning cannot be dealt with in this way, since the orientation between different pairs

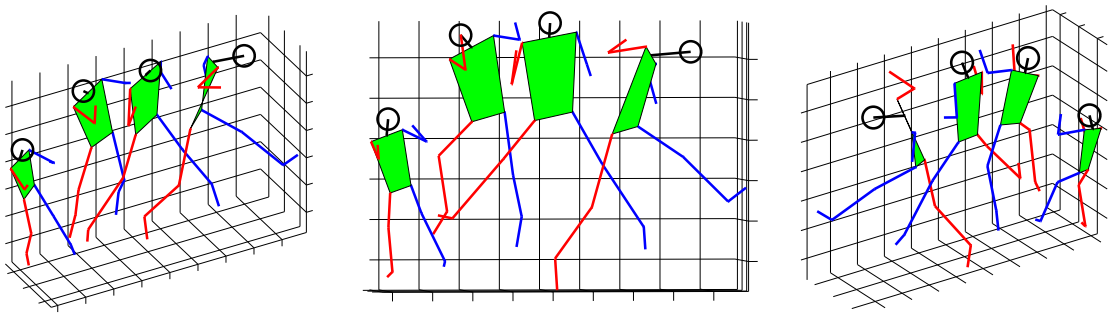


Figure 6.14: Metric reconstruction of the football sequence assuming fixed cameras.

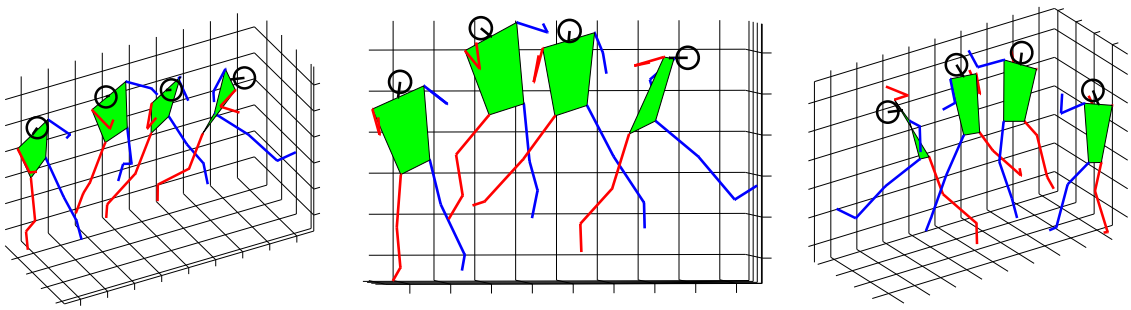


Figure 6.15: Fixed length body segment reconstruction of the football sequence assuming fixed cameras.

of cameras is not recovered. This would require some static scene object to be visible.

In the changing camera case, the rigid link constraint can be applied in two contexts. Firstly, symmetry constraints that the right and left arm and leg segments are the same length are valid in each reconstruction. There are thus at most four constraints on each $\Omega_\infty^{(j)}$ of the form (6.5).

Secondly, the rigid link constraint may be applied across reconstructions:

$$(\mathbf{D}^j)^\intercal \Omega_\infty^{(j)} (\mathbf{D}^j) - (\mathbf{D}^k)^\intercal \Omega_\infty^{(k)} (\mathbf{D}^k) = 0$$

where $\mathbf{D}^j$ and $\mathbf{D}^k$ represent the same rigid body segment in reconstructions j and k. This leads to a linear constraint on the vector of elements of both $\Omega_\infty^{(j)}$ and $\Omega_\infty^{(k)}$

$$\begin{aligned} (D_1^j)^2\Omega_1^{(j)} + 2(D_1^j D_2^j)\Omega_2^{(j)} + (D_2^j)^2\Omega_3^{(j)} + 2(D_1^j D_3^j)\Omega_4^{(j)} \\ +2(D_2^j D_3^j)\Omega_5^{(j)} + (D_3^j)^2\Omega_6^{(j)} - (D_1^k)^2\Omega_1^{(k)} + 2(D_1^k D_2^k)\Omega_2^{(k)} \\ +(D_2^k)^2\Omega_3^{(k)} + 2(D_1^k D_3^k)\Omega_4^{(k)} + 2(D_2^k D_3^k)\Omega_5^{(k)} + (D_3^k)^2\Omega_6^{(k)} \quad = \quad 0 \end{aligned} \tag{6.13}$$

The constraint can be written in vector form by constructing a vector of the elements of all the $\Omega_\infty^{(j)}$'s

$$\hat{\mathbf{\Omega}}_{\mathsf{v}} = (\Omega_1^{(1)},\ \Omega_2^{(1)},\ \Omega_3^{(1)},\ \Omega_4^{(1)},\ \ldots,\ \Omega_4^{(m)},\ \Omega_5^{(m)},\ \Omega_6^{(m)})^\intercal$$

and writing the coefficients of these elements in (6.13) as

$$\mathbf{c}^{(jk)} = (0,\ \ldots,\ 0,\ c_1^{(j)},\ c_2^{(j)},\ \ldots,\ c_6^{(j)},\ 0,\ \ldots,\ 0,\ c_1^{(k)},\ c_2^{(k)},\ \ldots c_6^{(k)},\ 0\ldots)^\intercal$$

The linear constraint in vector notation is then

$$(\mathbf{c}^{(jk)})^\intercal \hat{\mathbf{\Omega}}_{\mathsf{v}} = 0 \tag{6.14}$$

Up to nine constraints of this form can be written for each pair of reconstructions from the arms, legs and hips. Each $\Omega_\infty^{(j)}$ is over constrained, and rectification of each construction can be computed from the SVD of a constraint matrix whose rows are $\mathbf{c}^{(jk)}$ vectors

$$\mathsf{C}\hat{\mathbf{\Omega}}_{\mathsf{v}} = 0$$

Note that the $\Omega_\infty^{(j)}$'s computed are *not* homogeneous matrices in this context. The set of absolute conics that are computed are such that the relative lengths of segments in the metric reconstructions are the same, and thus *include* a scale factor for each reconstruction. In fact it is the vector of elements of all the absolute conics in the set, $\hat{\mathbf{\Omega}}_{\mathsf{v}}$, that is homogeneous. The common scale of this vector is a global scale factor applicable to all the reconstructions.

In practice, since there are far fewer constraints on each rectification than is the case for the single rectification with fixed cameras, it is useful to introduce camera parameters. Assuming that the cameras may be modeled as scaled orthographies, four equations on the dual of each absolute conic $(\Omega_\infty^{(j)})^{-1}$ can be obtained from (6.10) and (6.11) for each camera pair. These constraints are linear, but since they are in the dual conic, they cannot be simply combined with the rigid link constraints.

A solution is to use an iterative method to minimize a cost function over all the rectification

affinities $\mathsf{U}^{(j)}$. The cost function has two parts:

$$\mathcal{C} = \mathcal{C}_{struc} + \mathcal{C}_{cam}$$

The first term expresses the orthogonality and equal scale of the camera rows for cameras

$$\begin{aligned} \mathsf{G}_{\mathsf{M}}^{(j)} &= (\mathbf{g}_{\mathsf{M}1}^{j},\ \mathbf{g}_{\mathsf{M}2}^{j})^{\mathsf{T}} = \mathsf{G}^{(j)}(\mathsf{U}^{(j)})^{-1} \\ \text{and} \quad \mathsf{G}_{\mathsf{M}}^{\prime(j)} &= (\mathbf{g}_{\mathsf{M}1}^{\prime j},\ \mathbf{g}_{\mathsf{M}2}^{\prime j})^{\mathsf{T}} = \mathsf{G}^{\prime(j)}(\mathsf{U}^{(j)})^{-1} \end{aligned}$$

and is thus

$$\mathcal{C}_{cam} = \sum_{j=1}^{m} (\|\mathbf{g}_{\mathsf{M}1}^{j}\| - \|\mathbf{g}_{\mathsf{M}2}^{j}\|)^2 + (\|\mathbf{g}_{\mathsf{M}1}^{\prime j}\| - \|\mathbf{g}_{\mathsf{M}2}^{\prime j}\|)^2 + ((\mathbf{g}_{\mathsf{M}1}^{j})^{\mathsf{T}}\mathbf{g}_{\mathsf{M}2}^{j})^2 + ((\mathbf{g}_{\mathsf{M}1}^{\prime j})^{\mathsf{T}}\mathbf{g}_{\mathsf{M}2}^{\prime j})^2$$

The second term, $\mathcal{C}_{struc}$, is composed of terms of the form

$$(\|\mathsf{U}^{(j)}\mathbf{D}^{j}\| - \|\mathsf{U}^{(k)}\mathbf{D}^{k}\|)^2$$

for all applicable rigid links in the structure. The minimization can be initialized by a linear solution of $\hat{\mathbf{\Omega}}_v$ or from the fixed camera method.

The set of resulting $\mathsf{U}^{(j)}$'s returns m metric reconstructions with a common scale factor. However, since each initial affine reconstruction is a general affine transformation away from world structure, each has its own unknown Euclidean rectification component. That is, there remains a set of unknown rotations and translations between metric reconstructions of individual frame pairs. But, since the scaling is global to all the reconstructions, the relative scaling between pairs of cameras is known. The effect of a zooming camera can thus be eliminated by scaling the image points in each sequence relative to the first camera. Assuming negligible rotation of the cameras through the sequence, the fixed camera method can now be applied to compute a metric reconstruction of the *entire* motion.

Returning to the football example, consider figure 6.16, a plot of the scale factors of the two cameras calibrated as above. The cameras appear to be zooming in unison, with a ratio of 1.84.

(The fixed ratio between the camera scale factors suggests that there is global control of the zoom on the cameras filming the match.) The selected image points can now be scaled proportionately to negate the effect of the changing scale factor, as in figure 6.17. Applying the fixed camera algorithm to the scaled image data yields the results in figures 6.18, 6.19 and 6.20.

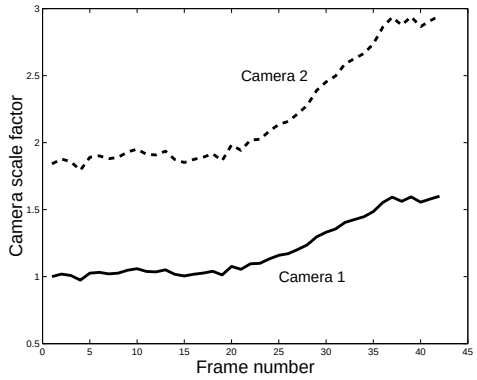


Figure 6.16: Scale factors of the two (scaled orthographic) cameras filming the football sequence. The scales are normalized to the first frame of camera 1.

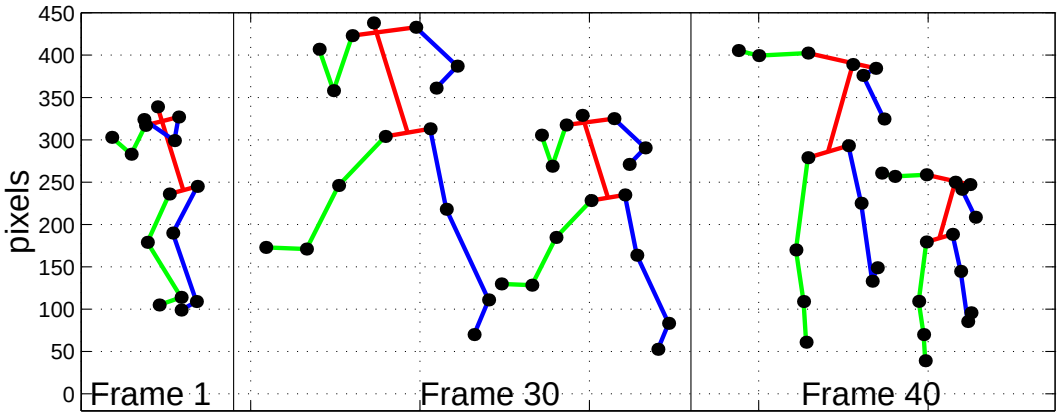


Figure 6.17: Zoom elimination in images from frames 1, 30 and 40 from the second camera in the football example. The selected image points for frame 1 appear on the left. In the middle and on the right (for frames 30 and 40) are large figures which are plots of the selected image points in those frames, and smaller figures obtained by scaling according to the recovered camera parameters.

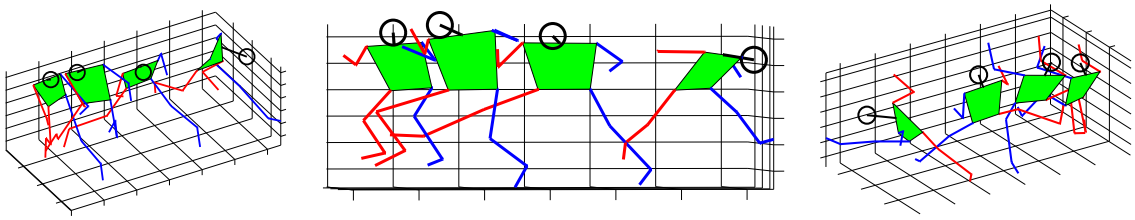


Figure 6.18: Affine reconstruction of the football sequence after zoom elimination. Note that the large scale increase seen in figure 6.13 is gone.

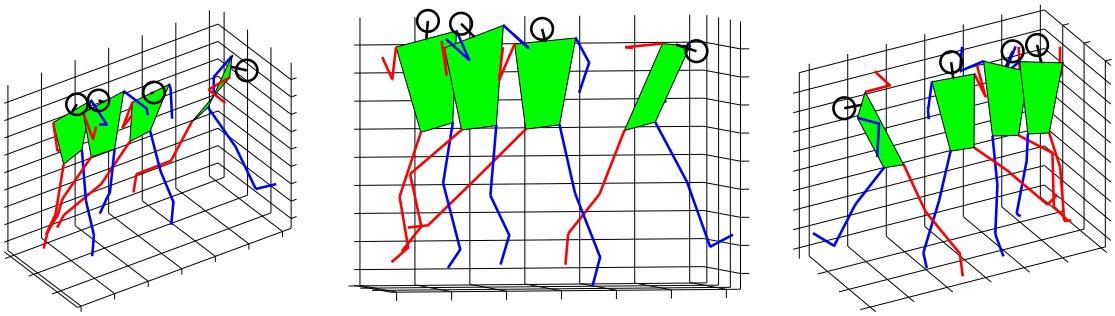


Figure 6.19: Final reconstruction of the football sequence.

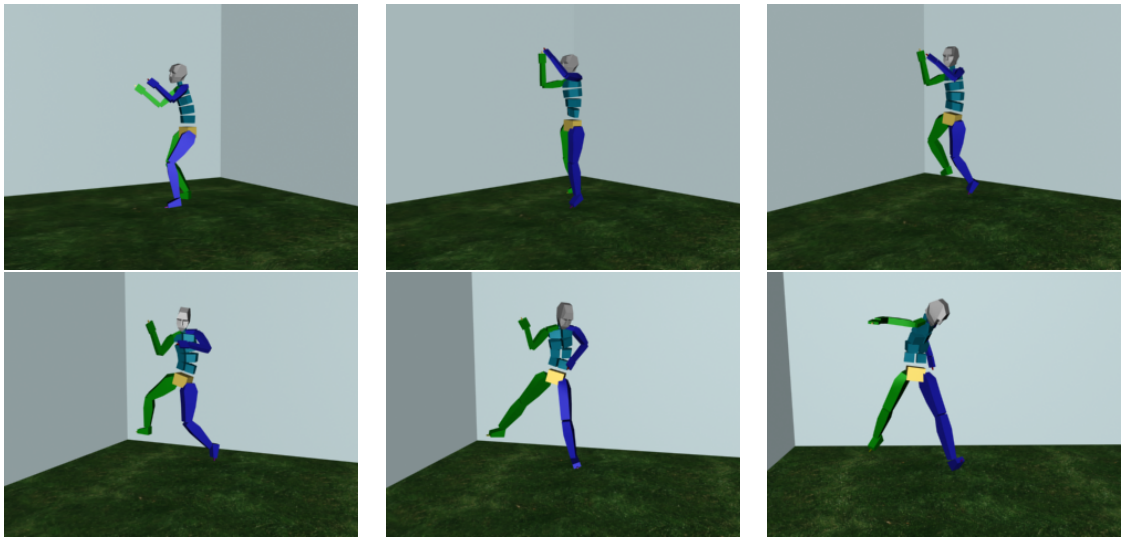


Figure 6.20: Six frames from an animation of the biped representation of the football example.

6.6 Summary

This chapter has presented a method of constraining the metric rectification of an affine reconstruction using the known relative lengths of line segments. The constraint has been applied to motion capture by modeling the human body as a rigid link articulated structure. The constant lengths of parts of the body as a person moves provide linear constraints on the absolute conic.

Using the affine camera model for cameras viewing sports events, an uncalibrated affine reconstruction is metric rectified by means of the rigid link constraint. The method has been demonstrated for fixed and zooming cameras to capture the motion of athletes from broadcast sports footage.

Chapter 7

Conclusion

In conclusion of the thesis, this chapter describes what are believed to be the novel contributions of the work and suggests areas of further research.

7.1 Novel contributions

The approach taken throughout the thesis is one of writing constraints on the geometric entities that specialize projective spaces. While the separate use of these entities has been reported in earlier work, combining them in one and two dimensions across single and multiple views is new. It has contributed to an understanding of the relationship between imaged planes and plane perspective distortion, structure recovery from one or more views, camera internal parameter constraints and auto-calibration.

- The metric plane geometry associated with the circular points constrains the image of the absolute conic, and thus the internal parameters of the camera. In turn, a known IAC constrains the imaged circular points of the plane, and defines them if the plane vanishing line is known. Thus the relationship between known internal camera parameters and imaged plane geometry is precisely defined by the circular points lying on the absolute conic. This has led to a novel object based calibration method of computing the IAC from a number of views of a plane with known metric structure. Other associated contributions are the techniques of extending the affine rectification of a plane to metric and MLE line intersection.

- If a camera has square pixels the image plane circular points are known. It was shown that these circular points provide two constraints on the IAC, just as any metric scene plane does. The image plane can thus be treated as a general scene plane in this sense.

- A single view of the vanishing points of two orthogonal directions places one constraint on the IAC. This shows how scene structure defines the IAC. The IAC, in turn, constrains scene reconstruction: the metric structure of the plane containing the two directions is fully defined by its circular points, which are the intersection of the vanishing line joining the vanishing points and the IAC.

- With the vanishing points of three orthogonal directions, three constraints on the IAC are available. If the camera has square pixels its circular points provide two constraints on the IAC and the five constraints in total fully define the internal calibration. This configuration has been used before, but the IAC based formulation allows a degeneracy analysis for a general configuration of three vanishing points and one metric plane. The analysis shows that when the vanishing line of the metric plane intersects one or two vertices of the triangle of vanishing points, there are only four independent constraints on the IAC.

- In two views, if the fundamental matrix is known, matched vanishing points in the images constrain the plane at infinity in a projective reconstruction. Three matched vanishing points thus allow rectification to an affine reconstruction. In addition, the three single view constraints on the IAC apply to each view individually. If the cameras have square pixels there are five constraints on the IAC which can be written in either view.

- The circular point parametrization of the square pixel property of a camera also defines an auto-calibration constraint. The circular points of either camera image plane can be transferred to the other view by the infinite homography. There are then seven constraints on the IAC. An extension of the single view degeneracy analysis shows that the number of constraints can be reduced to six or five, depending on the intersection of the vanishing points with the image plane and transferred image plane ideal lines.

- The case of the two cameras having identical internal parameters is similarly analysed. The

analysis is based on the geometric representation of the four constraints on the IAC from its invariance to transfer by the infinite homography. It shows that the seven constraints obtained from invariance and three vanishing points can be reduced to five or four.

- The final new result is the constraint on the absolute conic from the known relative lengths of a pair of line segments and its application in a motion capture system. The constraint is applied to the rigid links of the human body to reconstruct a motion sequence imaged by a pair of affine cameras. With fixed cameras, the entire motion is treated as a single static affine reconstruction of postures from the motion. Linear constraints on the absolute conic then allow metric rectification. If the cameras are zooming, constraints of the absolute conics in a set of reconstructions are written from rigid links. Including auto-calibration constraints on scaled orthographic cameras, the cameras and structure can be recovered.

7.2 Further research

Comments on areas of future work are divided into three categories: automation of data selection, error analysis and final section on possible extensions of the geometric constraints.

7.2.1 Automation

A common, and valid, criticism of the results shown here is that they depend on manual selection of data. User interaction to specify the geometry supplying constraints has been required for all the examples that have appeared. While such interaction (usually) provides data free of outliers, automation is desirable.

Many of the methods rely on the computation of vanishing points. There is a large body of literature on finding vanishing points automatically in structured environments, in which techniques as diverse as the Hough Transform[15, 108] and the Bayesian inference method of Coughlan and Yuille's *Manhattan World*[18] have been employed. Rother[86] computes three vanishing points specifically for camera calibration, and uses the the expected camera structure to test constraints of mutual orthogonality of the points. The planar grouping method of Schaffalitzky and Zisserman[90]

explicitly uses the geometry of repeated elements to compute vanishing points and lines. It also includes a method of finding the parameters specifying a regular grid that may be well suited to detecting features on calibration objects.

Any of these approaches should, however, be used with care. Many real world scenes contain geometry with three dominant directions, but often also have structure aligned closely but not perfectly to these directions. Adjacent buildings and roads, for example, are often not precisely parallel. Facade elements on old wooden buildings are frequently warped or not constructed with perfect orthogonality and parallelism in the first place. Even some classic architecture, notably the Taj Mahal[1], appears to fit the cuboid model, but is actually not constructed with its sides orthogonal and parallel. Experience with the two view measurement system described in chapter 5 indicates that even human selection of data is sometimes erroneous, and automatic methods are naturally prone to including such geometry.

Manual selection of joints for motion capture in a pair of sequences is particularly tedious. As a system, motion capture would be greatly improved by automated tracking to replace, or complement, the manual process. Fully automated tracking is unlikely to be completely successful given the quality of some digitised broadcast video. However, even partial tracking in conjunction with operator interaction will significantly speed up the process. There is a considerable amount of active research in the area, including most of the literature cited in chapter 6. It is feasible that simultaneous calibration from the rigid link constraint and tracking could be implemented for uncalibrated motion capture systems. The rigid link constraint could also be applied to marker data, where reflective markers are attached to the body to facilitate tracking.

7.2.2 Error analysis

First order error analysis has been performed for some of the algorithms described: the uncertainty in vanishing point computation and the camera parameters found using planes and three vanishing points have been quantified. One area requiring further work is extending the uncertainty in camera parameters to measurements computed with the aid of these parameters, such as the relative

[1]Discovering this fact while on a trip to India, some colleagues concluded that asking for a camera calibration from an image of the Taj Mahal would be a fine practical joke.

orientation of planes and the metric rectification parameters.

In particular, the two view measurement system requires that uncertainty in the computed 3D co-ordinates and distances should be supplied to the user, along the lines of the single view measurement system developed by Criminisi[19]. This means extending the single view analysis to two views, also taking into account uncertainty in the epipolar geometry. The motion capture system would similarly benefit from an analysis of sources of error and its propagation through to the reconstructed sequence.

The error analysis uses an assumption of Gaussian noise in image data. In the case of line intersection, the distribution error on endpoints of detected line segments is taken to be Gaussian. This determines the error in vanishing points and then in camera parameters. The validity of this assumption is questionable. Certainly, more work on the nature of the noise in image measurements of various types should be done, and incorporated into the error analysis.

The use of soft constraints on camera internal parameters results, in over constrained cases, in cameras where the constraints are not satisfied exactly. In the combination of conjugacy, square pixel and principal point constraints in chapter 4, the homogeneous constraints are written directly from the geometry. The resulting IAC, however, depends on the relative scaling of the rows of the coefficient matrix. A statistically rigorous application of priors such as the known principal point should consider the relative weighting of these constraints, based on their uncertainty. An investigation into the nature of this uncertainty on priors on camera parameters is thus warranted.

7.2.3 Geometry

An obvious extension of vanishing point based calibration is application to larger numbers of views. The two view calibration method using matched vanishing points of orthogonal directions can provide a good initialisation for subsequent auto-calibration bundle adjustment such as those of Pollefeys *et al*[79] and Nistér[75]. Inclusion of vanishing points is aided by the fact that the points can be computed in widely separated views without the same components of the structure being visible. The vanishing point of the same world direction can, for example, be computed from the facades of opposite sides of a building.

In addition, back-projecting the vanishing points provides a convenient parametrization of the absolute disk quadric. If the three points are $\mathbf{V}_1$, $\mathbf{V}_2$ and $\mathbf{V}_3$, Q_∞ can be written as a linear combination of three rank one degenerate dual quadrics:

$$\mathsf{Q}_\infty = k_1\mathbf{V}_1\mathbf{V}_1^\mathsf{T} + k_2\mathbf{V}_2\mathbf{V}_2^\mathsf{T} + k_3\mathbf{V}_3\mathbf{V}_3^\mathsf{T}$$

This parametrization is at most rank three, and thus explicitly satisfies the rank constraint on Q_∞. It is an over parametrization, requiring 15 degrees of freedom if the three points are homogeneous (and they should be, since by definition they may well become ideal in the computation).

In the motion capture method, a partial solution to the occlusion problem can be found when a section of semi-occluded frames can be omitted from a sequence, but the body can still be reconstructed from the remaining frames. An occluded point, now a *known* distance from an unoccluded 3D point, can lie anywhere on a sphere centred on the unoccluded point. If the occluded point is visible in one view, the back-projected ray defined by the image point intersects the sphere in two points, one of which is the occluded 3D point.

Bundle adjustment is also desirable for motion reconstruction to fit a truly rigid linked model to the image data. This is a complex non-linear optimization problem, and should use a kinematic model with appropriate degrees of freedom for various joints. It thus has a large number of degrees of freedom.

The relative length constraint implementation uses an affine camera model to generate an initial affine reconstruction. For some types of sports broadcasts, with cameras far from the players, this is a reasonable assumption. It is not, however, essential. An affine reconstruction can also be constructed from an initial projective reconstruction. The vanishing points of parallel lines bordering the field of play, for example, could be used to constrain the infinite homography. Orthogonal and parallel lines on the field also provide calibration information, as demonstrated by Reid and Zisserman[84] in their analysis of the famous football goal that wasn't.

Alternatively, the relative length constraint might be applied directly to the projective reconstruction, although it is no longer linear. The metric length of a line segment in a projective reconstruction can be written in terms of the elements of the metric rectifying homography N. The

metric rectification of endpoint $\mathbf{X}$, say, is $\mathsf{N}\mathbf{X}$, which must then be converted to inhomogeneous co-ordinates before a difference vector can be written. This is where the non-linearity emerges. Solution of the resulting system of non-linear equations must be initialized. One possibility for doing so is to make an initial affine camera assumption to obtain an estimate of metric structure. The transformation between the metric estimate and the projective reconstruction would then serve as an initial estimate of N.

Additional constraints of the type developed for plane rectification could also be applied in 3D. The constraint from equal unknown angles between pairs of lines, for example, can constrain the absolute conic.

Finally, the relative length constraint is a 3D constraint on Ω_∞, while most of the other constraints that have been used are image constraints on ω. They can still be combined, however. The homography between the image plane and the plane at infinity in a reconstruction is the left three by three sub-matrix of the projection matrix. For a canonical camera projection matrix $(\mathsf{I} \mid \mathbf{0})$, this homography is the identity and

$$\omega = \Omega_\infty.$$

Constraints can thus be transferred from one plane, and one conic, to the other.

Appendix A

The Jacobian of a 3×3 Cholesky decomposition

The internal camera parameter matrix K is typically recovered from the image of the absolute conic ω by Cholesky decomposition. Writing the elements of K as a vector $\mathbf{k}$, the terms of the Jacobian of the Cholesky decomposition $\mathsf{J}_{\mathbf{k}}^{\mathsf{T}} = \frac{\partial \mathbf{k}}{\partial \boldsymbol{\omega}_{\mathsf{v}}}$ are

$$
\begin{aligned}
\mathsf{J}_{\mathbf{k}11} &= \frac{-\omega_6\omega_1^2\omega_3^2 + 2\omega_6\omega_1\omega_3\omega_2^2 - \omega_6\omega_2^4 + 2\omega_4^2\omega_3^2\omega_1 - \omega_4^2\omega_3\omega_2^2 + \omega_5^2\omega_1^2\omega_3 - 4\omega_5\omega_4\omega_2\omega_3\omega_1 + 2\omega_5\omega_4\omega_2^3}{2\omega_1^{3/2}(\omega_3\omega_1 - \omega_2^2)^{3/2}(\omega_6\omega_1\omega_3 - \omega_6\omega_2^2 - \omega_4^2\omega_3 - \omega_5^2\omega_1 + 2\omega_5\omega_4\omega_2)^{1/2}} \\
J_{\mathbf{k}11} &= \frac{(\omega_5\omega_1 - \omega_2\omega_4)(-\omega_2\omega_5 + \omega_4\omega_3)}{\omega_1^{1/2}(\omega_6\omega_1\omega_3 - \omega_6\omega_2^2 - \omega_4^2\omega_3 - \omega_5^2\omega_1 + 2\omega_5\omega_4\omega_2)^{1/2}(\omega_3\omega_1 - \omega_2^2)^{3/2}} \\
J_{\mathbf{k}13} &= \frac{(-\omega_5\omega_1 + \omega_2\omega_4)^2}{2\omega_1^{1/2}(\omega_6\omega_1\omega_3 - \omega_6\omega_2^2 - \omega_4^2\omega_3 - \omega_5^2\omega_1 + 2\omega_5\omega_4\omega_2)^{1/2}(\omega_3\omega_1 - \omega_2^2)^{3/2}} \\
J_{\mathbf{k}14} &= \frac{\omega_2\omega_5 - \omega_4\omega_3}{\omega_1^{1/2}(\omega_6\omega_1\omega_3 - \omega_6\omega_2^2 - \omega_4^2\omega_3 - \omega_5^2\omega_1 + 2\omega_5\omega_4\omega_2)^{1/2}(\omega_3\omega_1 - \omega_2^2)^{1/2}} \\
J_{\mathbf{k}15} &= \frac{-\omega_5\omega_1 + \omega_4\omega_2}{\omega_1^{1/2}(\omega_6\omega_1\omega_3 - \omega_6\omega_2^2 - \omega_4^2\omega_3 - \omega_5^2\omega_1 + 2\omega_5\omega_4\omega_2)^{1/2}(\omega_3\omega_1 - \omega_2^2)^{1/2}} \\
J_{\mathbf{k}16} &= \frac{(\omega_3\omega_1 - \omega_2^2)^{1/2}}{2\omega_1^{1/2}(\omega_6\omega_1\omega_3 - \omega_6\omega_2^2 - \omega_4^2\omega_3 - \omega_5^2\omega_1 + 2\omega_5\omega_4\omega_2)^{1/2}}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{J}_{\mathbf{k}21} &= \frac{-\omega_2(-2\omega_1^2\omega_6\omega_3^2 + 3\omega_1\omega_6\omega_3\omega_2^2 + 3\omega_3^2\omega_1\omega_4^2 + 2\omega_1^2\omega_5^2\omega_3 - 6\omega_3\omega_1\omega_5\omega_4\omega_2 - \omega_2^4\omega_6 - \omega_2^2\omega_4^2\omega_3 + 2\omega_2^3\omega_5\omega_4)}{2\omega_1^{3/2}(\omega_3\omega_1 - \omega_2^2)^2(\omega_6\omega_1\omega_3 - \omega_6\omega_2^2 - \omega_4^2\omega_3 - \omega_5^2\omega_1 + 2\omega_5\omega_4\omega_2)^{1/2}} \\
\mathrm{J}_{\mathbf{k}22} &= \frac{-\omega_1^2\omega_6\omega_3^2 + \omega_1\omega_6\omega_3\omega_2^2 + \omega_3^2\omega_1\omega_4^2 + \omega_1^2\omega_5^2\omega_3 - 3\omega_3\omega_1\omega_5\omega_4\omega_2 + \omega_2^2\omega_4^2\omega_3 + \omega_1\omega_5^2\omega_2^2 - \omega_2^3\omega_5\omega_4}{(\omega_3\omega_1 - \omega_2^2)^2\omega_1^{1/2}(\omega_6\omega_1\omega_3 - \omega_6\omega_2^2 - \omega_4^2\omega_3 - \omega_5^2\omega_1 + 2\omega_5\omega_4\omega_2)^{1/2}} \\
\mathrm{J}_{\mathbf{k}23} &= \frac{-\omega_2(-\omega_6\omega_1^2\omega_3 + \omega_1\omega_6\omega_2^2 + \omega_4^2\omega_3\omega_1 + 2\omega_1^2\omega_5^2 - 4\omega_1\omega_5\omega_4\omega_2 + \omega_2^2\omega_4^2)}{2(\omega_3\omega_1 - \omega_2^2)^2\omega_1^{1/2}(\omega_6\omega_1\omega_3 - \omega_6\omega_2^2 - \omega_4^2\omega_3 - \omega_5^2\omega_1 + 2\omega_5\omega_4\omega_2)^{1/2}} \\
\mathrm{J}_{\mathbf{k}24} &= \frac{\omega_2(-\omega_2\omega_5 + \omega_4\omega_3)}{\omega_1^{1/2}(\omega_3\omega_1 - \omega_2^2)(\omega_6\omega_1\omega_3 - \omega_6\omega_2^2 - \omega_4^2\omega_3 - \omega_5^2\omega_1 + 2\omega_5\omega_4\omega_2)^{1/2}} \\
\mathrm{J}_{\mathbf{k}25} &= \frac{-\omega_2(-\omega_5\omega_1 + \omega_4\omega_2)}{\omega_1^{1/2}(\omega_3\omega_1 - \omega_2^2)(\omega_6\omega_1\omega_3 - \omega_6\omega_2^2 - \omega_4^2\omega_3 - \omega_5^2\omega_1 + 2\omega_5\omega_4\omega_2)^{1/2}} \\
\mathrm{J}_{\mathbf{k}26} &= \frac{-\omega_2}{2\omega_1^{1/2}(\omega_6\omega_1\omega_3 - \omega_6\omega_2^2 - \omega_4^2\omega_3 - \omega_5^2\omega_1 + 2\omega_5\omega_4\omega_2)^{1/2}}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{J}_{\mathbf{k}31} &= \frac{(-\omega_1\omega_6\omega_3\omega_2^2 + \omega_3^2\omega_1\omega_4^2 - 2\omega_3\omega_1\omega_5\omega_4\omega_2 + \omega_2^4\omega_6 + \omega_2^2\omega_4^2\omega_3 + 2\omega_1\omega_5^2\omega_2^2 - 2\omega_2^3\omega_5\omega_4)}{2(\omega_3\omega_1 - \omega_2^2)^2\omega_1^{1/2}(\omega_6\omega_1\omega_3 - \omega_6\omega_2^2 - \omega_4^2\omega_3 - \omega_5^2\omega_1 + 2\omega_5\omega_4\omega_2)^{1/2}} \\
\mathrm{J}_{\mathbf{k}32} &= \frac{-\omega_1^{1/2}(-\omega_6\omega_2\omega_3\omega_1 + \omega_6\omega_2^3 + 2\omega_2\omega_4^2\omega_3 + 2\omega_5^2\omega_1\omega_2 - 3\omega_5\omega_4\omega_2^2 - \omega_5\omega_4\omega_3\omega_1)}{(\omega_3\omega_1 - \omega_2^2)^2(\omega_6\omega_1\omega_3 - \omega_6\omega_2^2 - \omega_4^2\omega_3 - \omega_5^2\omega_1 + 2\omega_5\omega_4\omega_2)^{1/2}} \\
\mathrm{J}_{\mathbf{k}33} &= \frac{\omega_1^{1/2}(-\omega_6\omega_1^2\omega_3 + \omega_6\omega_1\omega_2^2 + \omega_4^2\omega_3\omega_1 + 2\omega_5^2\omega_1^2 - 4\omega_1\omega_5\omega_4\omega_2 + \omega_4^2\omega_2^2)}{2(\omega_3\omega_1 - \omega_2^2)^2(\omega_6\omega_1\omega_3 - \omega_6\omega_2^2 - \omega_4^2\omega_3 - \omega_5^2\omega_1 + 2\omega_5\omega_4\omega_2)^{1/2}} \\
\mathrm{J}_{\mathbf{k}34} &= \frac{-\omega_1^{1/2}(-\omega_2\omega_5 + \omega_4\omega_3)}{(\omega_3\omega_1 - \omega_2^2)(\omega_6\omega_1\omega_3 - \omega_6\omega_2^2 - \omega_4^2\omega_3 - \omega_5^2\omega_1 + 2\omega_5\omega_4\omega_2)^{1/2}} \\
\mathrm{J}_{\mathbf{k}35} &= \frac{\omega_1^{1/2}(-\omega_5\omega_1 + \omega_4\omega_2)}{(\omega_3\omega_1 - \omega_2^2)(\omega_6\omega_1\omega_3 - \omega_6\omega_2^2 - \omega_4^2\omega_3 - \omega_5^2\omega_1 + 2\omega_5\omega_4\omega_2)^{1/2}} \\
\mathrm{J}_{\mathbf{k}36} &= \frac{\omega_1^{1/2}}{2(\omega_6\omega_1\omega_3 - \omega_6\omega_2^2 - \omega_4^2\omega_3 - \omega_5^2\omega_1 + 2\omega_5\omega_4\omega_2)^{1/2}}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{J}_{\mathbf{k}41} &= \frac{-\omega_2\omega_5 + \omega_4\omega_3}{(\omega_3\omega_1 - \omega_2^2)^2\omega_3} \\
\mathrm{J}_{\mathbf{k}42} &= \frac{\omega_5\omega_3\omega_1 + \omega_5\omega_2^2 - 2\omega_2\omega_4\omega_3}{(\omega_3\omega_1 - \omega_2^2)^2} \\
\mathrm{J}_{\mathbf{k}43} &= \frac{\omega_2(\omega_2\omega_4 - \omega_5\omega_1)}{(\omega_3\omega_1 - \omega_2^2)^2} \\
\mathrm{J}_{\mathbf{k}44} &= \frac{-\omega_3}{\omega_3\omega_1 - \omega_2^2} \\
\mathrm{J}_{\mathbf{k}45} &= \frac{\omega_2}{\omega_3\omega_1 - \omega_2^2} \\
\mathrm{J}_{\mathbf{k}46} &= 0
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{J}_{\mathbf{k}5\,1} &= \frac{\omega_2(\omega_2\omega_5 - \omega_4\omega_3)}{(\omega_3\omega_1 - \omega_2^2)^2} \\
\mathrm{J}_{\mathbf{k}5\,2} &= \frac{\omega_4\omega_3\omega_1 + \omega_4\omega_2^2 - 2\omega_1\omega_2\omega_5}{(\omega_3\omega_1 - \omega_2^2)^2} \\
\mathrm{J}_{\mathbf{k}5\,3} &= \frac{\omega_1(-\omega_2\omega_4 + \omega_5\omega_1)}{(\omega_3\omega_1 - \omega_2^2)^2} \\
\mathrm{J}_{\mathbf{k}5\,4} &= \frac{\omega_2}{\omega_3\omega_1 - \omega_2^2} \\
\mathrm{J}_{\mathbf{k}5\,5} &= \frac{\omega_1}{\omega_3\omega_1 - \omega_2^2} \\
\mathrm{J}_{\mathbf{k}5\,6} &= 0
\end{aligned}
$$

Bibliography

[1] Y. I. Abdel-Aziz and H. M. Karara. Direct linear transformation into object space coordinates in close-range photogrammetry. In *Symposium on Close-Range Photogrammetry*, 1971.

[2] M. Armstrong, A. Zisserman, and P. Beardsley. Euclidean reconstruction from uncalibrated images. In *Proc. British Machine Vision Conference*, pages 509–518, 1994.

[3] M. Armstrong, A. Zisserman, and R. Hartley. Self-calibration from image triplets. In *Proc. European Conference on Computer Vision*, LNCS 1064/5, pages 3–16. Springer-Verlag, 1996.

[4] P. Beardsley, P. Torr, and A. Zisserman. 3D model acquisition from extended image sequences. In *Proc. European Conference on Computer Vision*, LNCS 1064/1065, pages 683–695. Springer-Verlag, 1996.

[5] P. Beardsley and A. Zisserman. Affine calibration of mobile vehicles. In Mohr, R. and Chengke, W., editors, *Europe-China workshop on Geometrical Modelling and Invariants for Computer Vision*, pages 214–221. Xidan University Press, Xi'an, China, 1995.

[6] P. A. Beardsley, A. Zisserman, and D. W. Murray. Sequential update of projective and affine structure from motion. *International Journal of Computer Vision*, 23(3):235–259, 1997.

[7] S. Bougnoux. From Projective to Euclidean space under any practical situation, a criticism of self-calibration. In *Proc. 6th International Conference on Computer Vision, Bombay, India*, pages 790–796, January 1998.

[8] J.M. Brady and A.L. Yuille. An extremum principal for shape from contour. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(3):288–301, May 1984.

[9] C. Bregler and J. Malik. Video motion capture. Technical Report UCB//CSD-97-973, University of California, Berkeley, 1997.

[10] J. F. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.

[11] B. Caprile and V. Torre. Using vanishing points for camera calibration. *International Journal of Computer Vision*, pages 127–140, 1990.

[12] R. Cipolla and E.G. Boyer. 3D model acquisition from uncalibrated images. In *Proc. IAPR Workshop on Machine Vision Applications, Chiba, Japan*, pages 559–568, November 1998.

[13] J. C. Clarke. First order error propagation: A primer. *Image and Vision Computing*, 1996.

[14] J. C. Clarke. Modelling uncertainty: A primer. Technical Report 2161/98, University of Oxford, Dept. Engineering Science, 1998.

[15] R. T. Collins and J. R. Beveridge. Matching perspective views of coplanar structures using projective unwarping and similarity matching. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 1993.

[16] R.T. Collins. A Bayesian anlysis of projective incidence. In J.L. Mundy and A. Zisserman, editors, *Proc. 2nd European-US Workshop on Invariance, Azores*, pages 151–163, 1993.

[17] The TargetJr Consortium. *http://www.targetjr.org*. 2000.

[18] J.M. Coughlan and A.L. Yuille. Manhattan world: Compass direction from a single image by bayesian inference. In *Proc. 7th International Conference on Computer Vision, Kerkyra, Greece*, pages 941–947, 1999.

[19] A. Criminisi. *Accurate Visual Metrology from Single and Multiple Uncalibrated Images*. PhD thesis, University of Oxford, Dept. Engineering Science, December 1999. D.Phil. thesis.

[20] A Criminisi, I. Reid, and A Zisserman. A plane measuring device. *Image and Vision Computing*, 17(8):625–634, 1999.

[21] A. Criminisi, I. Reid, and A. Zisserman. Single view metrology. In *Proc. 7th International Conference on Computer Vision, Kerkyra, Greece*, pages 434–442, September 1999.

[22] K. Daniilidis and J. Ernst. Active intrinsic calibration using vanishing points. *Pattern Recognition Letters*, 17:1179–1189, 1996.

[23] L. de Agapito, R. Hartley, and E. Hayman. Linear calibration of a rotating and zooming camera. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition, Fort Collins, Colorado*, 1999.

[24] L. de Agapito, E. Hayman, and I. Reid. Self-calibration of a rotating camera with varying intrinsic parameters. In *Proc. 9th British Machine Vision Conference, Southampton*, 1998.

[25] P. E. Debevec, C. J. Taylor, and J. Malik. Modeling and rendering architecture from photographs: A hybrid geometry- and image- based approach. In *Proceedings, ACM SIGGRAPH*, pages 11–20, 1996.

[26] J. Deutscher, A. Blake, and I. Reid. Articulated body motion capture by annealed particle filtering. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition, Hilton Head Island, South Carolina*, 2000.

[27] F. Devernay and O. D. Faugeras. Automatic calibration and removal of distortion from scenes of structured environments. In *SPIE*, volume 2567, San Diego, CA, July 1995.

[28] F. Devernay and O. D. Faugeras. From projective to euclidean reconstruction. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 264–269, 1996.

[29] O. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In *Proc. European Conference on Computer Vision*, LNCS 588, pages 563–578. Springer-Verlag, 1992.

[30] O. Faugeras and T. Papadopoulo. Grassman-Cayley algebra for modelling systems of cameras and the algebraic equations of the manifold of trifocal tensors. *Philosophical Transactions of the Royal Society of London, SERIES A*, 356(1740):1123–1152, 1998.

[31] O. D. Faugeras. *Three-Dimensional Computer Vision: a Geometric Viewpoint.* MIT Press, 1993.

[32] O. D. Faugeras, S. Laveau, L. Robert, G. Csurka, and C. Zeller. 3-D reconstruction of urban scenes from sequences of images. Tech. report, INRIA, 1995.

[33] O. D. Faugeras, Q. Luong, and S. Maybank. Camera self-calibration: Theory and experiments. In *Proc. European Conference on Computer Vision*, LNCS 588, pages 321–334. Springer-Verlag, 1992.

[34] O. D. Faugeras and S. J. Maybank. Motion from point matches: Multiplicity of solutions. *International Journal of Computer Vision*, 4:225–246, 1990.

[35] A. W. Fitzgibbon and A. Zisserman. Automatic camera recovery for closed or open image sequences. In *Proc. European Conference on Computer Vision*, pages 311–326. Springer-Verlag, June 1998.

[36] J. D. Foley, A. Van Dam, S. K. Feiner, and J. F. Hughes. *Computer Graphics: Principles and Practice.* Addison-Wesley, 1990.

[37] S. Ganapathy. Decomposition of transformation matrices for robot vision. In *Proc. International Conference on Robotics and Automation*, pages 130–139, 1984.

[38] D.M. Gavrila. The visual analysis of human movement: A survey. *Computer Vision and Image Understanding*, 73(1), 1999.

[39] G. Golub and C. Van Loan. *Matrix Computations.* John Hopkins University Press, 1983.

[40] G. H. Golub and C.F. Van Loan. *Matrix Computations.* The John Hopkins University Press, Baltimore, MD, second edition, 1989.

[41] G. Gracie. Analytical photogrammetry applied to single terrestrial photograph mensuration. In *XIth International Conference of Photogrammetry, Lausanne, Switzerland*, July 1968.

[42] C. J. Harris and M. Stephens. A combined corner and edge detector. In *Proc. 4th Alvey Vision Conference, Manchester*, pages 147–151, 1988.

[43] R. I. Hartley. Estimation of relative camera positions for uncalibrated cameras. In *Proc. European Conference on Computer Vision*, LNCS 588, pages 579–587. Springer-Verlag, 1992.

[44] R. I. Hartley. Euclidean reconstruction from uncalibrated views. In J.L. Mundy, A. Zisserman, and D. Forsyth, editors, *Proc. 2nd European-US Workshop on Invariance, Azores*, pages 187–202, 1993.

[45] R. I. Hartley. Self-calibration from multiple views with a rotating camera. In *Proc. European Conference on Computer Vision*, LNCS 800/801, pages 471–478. Springer-Verlag, 1994.

[46] R. I. Hartley. Kruppa's equations derived from the fundamental matrix. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(2):133–135, 1997.

[47] R. I. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 1992.

[48] R. I. Hartley and P. Sturm. Triangulation. In *Proc. Conference Computer Analysis of Images and Patterns*, Prague, Czech Republic, 1995.

[49] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision.* Cambridge University Press, ISBN: 0521623049, 2000.

[50] R.I. Hartley, E. Hayman, and L. de Agapito. Camera calibration and the search for infinity. In *Proc. 7th International Conference on Computer Vision, Kerkyra, Greece*, pages 510–517, 1999.

[51] A. Heyden and K. Åström. Euclidean reconstruction from constant intrinsic parameters. In *Proc. International Conference on Pattern Recognition*, 1996.

[52] A. Heyden and K. Åström. Euclidean reconstruction from image sequences with varying and unknown focal length and principal point. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 1997.

[53] Berthold K. P. Horn and Michael J. Brooks. *Shape from Shading.* The MIT Press, Cambridge, MA, 1989.

[54] Y. Horry, K. Anjyo, and K. Arai. Tour into the picture: Using a spidery mesh interface to make animation from a single image. In *Proceedings of the ACM SIGGRAPH Conference on Computer Graphics*, pages 225–232, 1997.

[55] W. H. Ittelson. *The Ames Demonstrations in Perception.* Princeton University Press, Princeton, New Jersey, 1952.

[56] K. Kanatani. *Geometric computation for machine vision.* Oxford University Press, Oxford, 1992.

[57] K. Kanatani. Statistical optimization for geometric computation: theory and practice. Technical report, AI Lab, Dept of Computer Science, Gunma University, 1995.

[58] G.E. Karras, P. Patias, and E. Petsa. Experiences with rectification of non-metric digital images when ground control is not available. In *CIPA XV International Symposium*, 1993.

[59] M. Kemp. *The Science of Art.* Yale University Press, New Haven and London, 1989.

[60] M. (editor) Kemp. *Leonardo on Painting.* Yale University Press, New Haven and London, 1989.

[61] J. J. Koenderink. Pictorial relief. *Philosophical Transactions of the Royal Society of London, SERIES A*, 356(1740):1071–1083, 1998.

[62] J. J. Koenderink and A. J. van Doorn. Affine structure from motion. *J. Opt. Soc. Am. A*, 8(2):377–385, 1991.

[63] E. Kreysig. *Advanced Engineering Mathematics.* John Wiley and Sons, New York, 1993.

[64] M.E. Leventon and W.T. Freeman. Bayesian estimation of 3-D human motion from an image sequence. Technical Report TR-98-06, MERL, 1998.

[65] D. Liebowitz and S. Carlsson. Uncalibrated motion capture exploiting articulated structure constraints. In *Proc. 8th International Conference on Computer Vision, Vancouver, Canada*, 2001.

[66] D. Liebowitz, A. Criminisi, and A. Zisserman. Creating architectural models from images. In *Proc. EuroGraphics*, volume 18, pages 39–50, September 1999.

[67] D. Liebowitz and A. Zisserman. Metric rectification for perspective images of planes. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 482–488, June 1998.

[68] D. Liebowitz and A. Zisserman. Combining scene and auto-calibration constraints. In *Proc. 7th International Conference on Computer Vision, Kerkyra, Greece*, September 1999.

[69] Q. Luong and T. Viéville. Canonic representations for the geometries of multiple projective views. Technical Report UCB/CSD-93-772, University of California, Berkeley, USA, 1993.

[70] Q. T. Luong and O. D. Faugeras. The fundamental matrix: theory, algorithms, and stability analysis. *International Journal of Computer Vision*, 17(1):43–76, 1996.

[71] Q.-T. Luong and T. Viéville. Canonic representations for the geometries of multiple projective views. In *Proc. 3rd European Conference on Computer Vision, Stockholm*, pages 589–599, May 1994.

[72] T. Moons, L. Van Gool, M. Van Diest, and A. Oosterlinck. Affine structure from perspective image pairs obtained by a translating camera. In J. L. Mundy, A. Zisserman, and D. Forsyth, editors, *Applications of invariance in computer vision*, pages 297–316. Springer-Verlag, 1994.

[73] J. Mundy and A. Zisserman. *Geometric Invariance in Computer Vision.* MIT Press, 1992.

[74] S. Nayar and S. Narasimhan. Vision in bad weather. In *Proc. 7th International Conference on Computer Vision, Kerkyra, Greece*, pages 820–827, September 1999.

[75] D. Nistér. *Automatic Dense Reconstruction from Uncalibrated Video Sequences.* PhD thesis, Kungl Tekniska Högskolan (Royal Institute of Technology), Sweden, 2001.

[76] T. Pajdla and V. Hlaváč. Camera calibration and Euclidean reconstruction from known observer translations. In *International Conference on Computer Vision and Pattern Recognition*, pages 421–426. IEEE, June 1998.

[77] S. E. Palmer. *Vision Science.* MIT Press, Cambridge, Massachusetts, 1999.

[78] C. Poelman and T. Kanade. A paraperspective factorization method for shape and motion recovery. In *Proc. 3rd European Conference on Computer Vision, Stockholm*, volume 2, pages 97–108, 1994.

[79] M. Pollefeys, R. Koch, and L. Van Gool. Self calibration and metric reconstruction in spite of varying and unknown internal camera parameters. In *Proc. 6th International Conference on Computer Vision, Bombay, India*, pages 90–96, 1998.

[80] M. Pollefeys and L. Van Gool. A stratified approach to metric self calibration. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 1997.

[81] M. Pollefeys, L. Van Gool, and A. Oosterlinck. The modulus constraint: a new constraint for self-calibration. In *Proc. International Conference on Pattern Recognition*, pages 31–42, 1996.

[82] W. Press, B. Flannery, S. Teukolsky, and W. Vetterling. *Numerical Recipes in C.* Cambridge University Press, 1988.

[83] L. Quan. Self-calibration of an affine camera from multiple views. *International Journal of Computer Vision*, 19(1):93–105, 1996.

[84] I. Reid and A. Zisserman. Goal-directed video metrology. In R. Cipolla and B. Buxton,

editors, *Proc. 4th European Conference on Computer Vision, LNCS 1065, Cambridge*, volume II, pages 647–658. Springer, April 1996.

[85] I. D. Reid and D. W. Murray. Active tracking of foveated feature clusters using affine structure. *International Journal of Computer Vision*, 18(1):41–60, 1996.

[86] C. Rother. A new approach for vanishing point detection in architectural environments. In *Proc. 11th British Machine Vision Conference, Bristol*, 2000.

[87] C. Rothwell, A. Zisserman, J. Mundy, and D. Forsyth. Efficient model library access by projectively invariant indexing functions. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 109–114, 1992.

[88] D.J. Saville and G.R. Wood. *Statistical Methods: The Geometric Approach*. Springer, 1991.

[89] F. Schaffalitzky. Direct solution of modulus constraints. In *Proc. Indian Conference on Computer Vision, Graphics and Image Processing*, pages 314–321, 2000.

[90] F. Schaffalitzky and A. Zisserman. Planar grouping for automatic detection of vanishing lines and points. *Image and Vision Computing*, 18:647–658, 2000.

[91] J.G. Semple and G.T. Kneebone. *Algebraic Projective Geometry*. Oxford University Press, 1979.

[92] L. Shapiro, A. Zisserman, and M. Brady. Motion from point matches using affine epipolar geometry. In *Proc. European Conference on Computer Vision*, LNCS 800/801. Springer-Verlag, 1994.

[93] L. S. Shapiro, A. Zisserman, and M. Brady. 3D motion recovery via affine epipolar geometry. *International Journal of Computer Vision*, 16(2):147–182, 1995.

[94] J. A. Shufelt. Performance and analysis of vanishing point detection techniques. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(3):282–288, March 1999.

[95] D. Sinclair, L. Paletta, and A. Pinz. Euclidean structure recovery through articulated motion. In *Proc. Scandinavian Conference on Image Analysis*, 1997.

[96] C. Slama. *Manual of Photogrammetry*. American Society of Photogrammetry, Falls Church, VA, USA, 4th edition, 1980.

[97] P. Sturm and S. Maybank. On plane based camera calibration: A general algorithm, singularities, applications. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 432–437, June 1999.

[98] P. Sturm and S. J. Maybank. A method for interactive 3D reconstruction of piecewise planar objects from single images. In *Proc. 10th British Machine Vision Conference, Nottingham*, 1999.

[99] R. Swaminathan and S.K. Nayar. Non-metric calibration of wide-angle lenses and polycameras. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition, Fort Collins, Colorado*, 1999.

[100] R. Szeliski. Image mosaicing for tele-reality applications. Technical report, Digital Equipment Corporation, Cambridge, USA, 1994.

[101] R. Szeliski and S. Heung-Yeung. Creating full view panoramic image mosaics and environment maps. In *Proceedings of the ACM SIGGRAPH Conference on Computer Graphics*, 1997.

[102] C.J. Taylor. Reconstruction of articulated objects from point correspondences in a single uncalibrated image. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, June 2000.

[103] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization approach. *International Journal of Computer Vision*, 9(2):137–154, November 1992.

[104] P. H. S. Torr and D. W. Murray. The development and comparison of robust methods for estimating the fundamental matrix. *International Journal of Computer Vision*, 24(3):271–300, 1997.

[105] W. Triggs. Auto-calibration and the absolute quadric. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 609–614, 1997.

[106] W. Triggs. Autocalibration from planar scenes. In *Proc. 5th European Conference on Computer Vision, Freiburg, Germany*, 1998.

[107] Y. R. Tsai. An efficient and accurate camera calibration technique for 3D machine vision. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 1986.

[108] T. Tuytelaars, L. Van Gool, M. Proesmans, and T. Moons. The cascaded Hough transform as an aid in aerial image interpretation. In *Proc. 6th International Conference on Computer Vision, Bombay, India*, pages 67–72, January 1998.

[109] J. A. Webb and J. K. Aggarwal. Structure from motion of rigid and jointed object. *Artificial Intelligence*, 19:107–130, 1983.

[110] D. Weinshall and C. Tomasi. Linear and incremental acquisition of invariant shape models from image sequences. In *Proc. 4th International Conference on Computer Vision, Berlin*, pages 675–682, Los Alamitos, CA, 1993. IEEE Computer Society Press.

[111] A. P. Witkin. Recovering surface shape and orientation from texture. *Artificial Intelligence*, 1981.

[112] C. Zeller. *Projective, Affine and Euclidean Calibration in Compute Vision and the Application of Three Dimensional Perception.* PhD thesis, RobotVis Group, INRIA Sophia-Antipolis, 1996.

[113] C. Zeller and O. D. Faugeras. Camera self-calibration from video sequences: the kruppa equations revisited. Technical Report 2793, INRIA, 1995.

[114] Z. Zhang. A flexible new technique for camera calibration. In *Proc. 7th International Conference on Computer Vision, Kerkyra, Greece*, September 1999.

[115] A. Zisserman. Notes on geometric invariance in vision. Tutorial, British Machine Vision Conference, 1992.

[116] A. Zisserman, P. Beardsley, and I. Reid. Metric calibration of a stereo rig. In *IEEE Workshop on Representation of Visual Scenes, Boston*, pages 93–100, 1995.

[117] A. Zisserman, D. Liebowitz, and M. Armstrong. Resolving ambiguities in auto-calibration. *Philosophical Transactions of the Royal Society of London, SERIES A*, 356(1740):1193–1211, 1998.