Projective Geometry
for
Image Analysis

A Tutorial given at ISPRS, Vienna, July 1996

Roger Mohr and Bill Triggs

GRAVIR, project MOVI
INRIA, 655 avenue de l'Europe
F-38330 Montbonnot St Martin
France

E-mail: {*Roger.Mohr,Bill.Triggs*}@*inrialpes.fr*
WWW: *http://www.inrialpes.fr/movi*

September 25, 1996

Contents

Chapter 1

Foreword and Motivation

Significant progress has recently been made by applying tools from classical projective and algebraic geometry to fundamental problems in computer vision. To some extent this work was foreshadowed by early mathematical photogrammetrists. However the modern approach has gone far beyond these early studies, particularly as regards our ability to deal with multiple images and unknown camera parameters, and on practical computational issues such as stability, robustness and precision. These new techniques are greatly extending the scope and flexibility of digital photogrammetric systems.

This tutorial provides a practical, applications-oriented introduction to the projective geometry needed to understand these new developments. No currently available textbook covers all of this material, although several existing texts consider parts of it. Kanatani's book [11] studies many computational and statistical aspects of computer vision in a projective framework. Faugeras [4] investigates the geometric aspects of 3D vision, including several of the projective results obtained by his team before 1993. The collections edited by Mundy, Zisserman and Forsyth [18, 19] summarize recent research on the applications of geometric invariants to computer vision: projective results are central to this programme.

Mathematical introductions to projective geometry can be found in many books. A standard text covering the necessary aspects of both projective and algebraic geometry is Semple and Kneebone [23]. Unfortunately this is currently out of print, however many scientific libraries have it and it is said to be reprinting soon.

Synopsis: Chapter 1 provides initial motivation and discusses the perspective model of image projection. Chapter 2 formally describes the basic properties of projective space. Chapter 3 considers projective invariants and the cross ratio. Chapter 4 compares the basic structures of projective, affine and Euclidean geometries and shows how to specialize from one to the other. Finally, chapter 5 considers the problems of 3D reconstruction using uncalibrated cameras and camera auto-calibration under various prior hypotheses.

How to read these notes

You should be familiar with elementary calculus and linear algebra. Exercises are provided for each chapter. You should at least skim through these: some just provide practice with the required computations, but others consider important problems and pitfalls and need to be addressed more carefully.

1.1 Intuitive Considerations About Perspective Projection

1.1.1 An Infinitely Strange Perspective

The study of projective geometry was initiated by the painters of the Italian Renaissance, who wanted to produce a convincing illusion of 3D depth in their architectural paintings. They made considerable use of vanishing points and derived several practically useful geometric constructions, for example to split a projected square into four equal sub-squares, or to find the projection of a parallelogram when the projection of two of its sides are known.

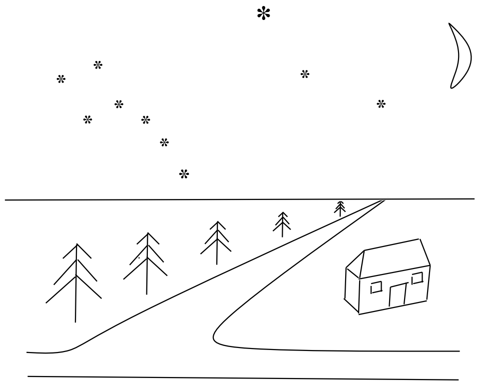


Figure 1.1: Landscape with horizon

Look at figure 1.1: the edges of the road are parallel lines in 3D space, but in the image they appear to converge as they recede towards the horizon. The line of the horizon is formed by the "infinitely distant points" or *vanishing directions* of the ground plane. Any pair of parallel, horizontal lines appears to meet at the point of the horizon corresponding to their common direction. This is true even if they lie at different heights above the ground plane. Moreover, any two horizontal planes appear to come together in the distance, and intersect in the horizon line or "line at infinity".

All of these "intersections at infinity" stay constant as the observer moves. The road always seems to disappear at the same point (direction) on the horizon, and the stars stay fixed as you walk along: lines of sight to infinitely distant points are always parallel, because they "(only) meet at infinity".

These simple examples show that our usual concepts of finite geometry have to be extended to handle phenomena "at infinity" that project to very finite locations in images.

1.1.2 Homogeneous Coordinates

How can we handle all this mathematically? — Every point in an image represents a possible line of sight of an incoming light ray: any 3D point along the ray projects to the same image point, so only the direction of the ray is relevant, not the distance of the point along it. In vision we need to represent this "celestial" or "visual sphere" of incoming ray directions. One way to do this is by their two image (*e.g.* pixel) coordinates (x, y). Another is by arbitrarily choosing some 3D point along each ray to represent the ray's direction. In this case we need three "homogeneous coordinates" instead of two "inhomogeneous" ones to represent each ray. This seems inefficient, but it has the significant advantage of making the image projection process much easier to deal with.

In detail, suppose that the camera is at the origin $(0, 0, 0)$. The ray represented by "homogeneous coordinates" (X, Y, T) is that passing through the 3D point (X, Y, T). The 3D point $\lambda \cdot (X, Y, T) = (\lambda X, \lambda Y, \lambda T)$ also lies on (represents) the same ray, so we have the rule that rescaling homogeneous coordinates makes no difference:

$$(X, Y, T) \sim \lambda(X, Y, T) = (\lambda X, \lambda Y, \lambda T)$$

If we suppose that the image plane of the camera is $T = 1$, the ray through pixel (x, y) can be represented homogeneously by the vector $(x, y, 1) \sim (xT, yT, T)$ for any depth $T \neq 0$. Hence, the homogeneous point vector (X, Y, T) with $T \neq 0$ corresponds to the inhomogeneous image point $(\frac{X}{T}, \frac{Y}{T})$ on the plane $T = 1$.

But what happens when $T = 0$? — $(X, Y, 0)$ is a valid 3D point that defines a perfectly normal optical ray, but this ray does not correspond to any finite pixel: it is parallel to the plane $T = 1$ and so has no finite intersection with it. Such rays or homogeneous vectors can no longer be interpreted as finite points of the standard 2D plane. However, they can be viewed as additional "ideal points" or limits as (x, y) recedes to infinity in a certain direction:

$$\lim_{T \to 0} (\frac{X}{T}, \frac{Y}{T}, 1) \sim \lim_{T \to 0} (X, Y, T) = (X, Y, 0)$$

We can add such ideal points to any 3D plane. In 2D images of the plane, the added points at infinity form the plane's "horizon". We can also play the same trick on the whole 3D space, representing 3D points by four homogeneous coordinates $(X, Y, Z, T) \sim (\lambda X, \lambda Y, \lambda Z, \lambda T) \sim (\frac{X}{T}, \frac{Y}{T}, \frac{Z}{T}, 1)$ and adding a "plane at infinity" $T = 0$ containing an "ideal point at infinity" for each 3D direction, represented by the homogeneous vector $(X, Y, Z, 0)$. This may seem unnecessarily abstract, but it turns out that 3D visual reconstruction is most naturally expressed in terms of such a "3D projective space", so the theory is well worth studying.

Line coordinates: The planar line with equation $ax + by + c = 0$ is represented in homogeneous coordinates by the homogeneous equation $(a, b, c) \cdot (X, Y, T) = aX + bY + cT = 0$. If the line vector (a, b, c) is $(0, 0, 1)$ we get the special "line" $T = 0$ which contains only ideal points and is called the *line at infinity*. Note that lines are represented homogeneously as 3 component vectors, just as points are. This is the first sign of a deep and powerful *projective duality* between points and lines.

Now consider an algebraic curve. The standard hyperbola has equation $xy = 1$. Substitute $x = \frac{X}{T}, y = \frac{Y}{T}$ and multiply out to get $XY = T^2$. This is homogeneous of degree 2. In fact, in homogeneous coordinates, any polynomial can be re-expressed as a homogeneous one. Notice that $(0, \lambda, 0)$ and $(\lambda, 0, 0)$ are valid solutions of $XY = T^2$: the homogeneous hyperbola crosses the $\vec{x}$ axis smoothly at $y = \infty$ and the $\vec{y}$ axis smoothly at $x = \infty$, and comes back on the other side (see fig. 1.2).

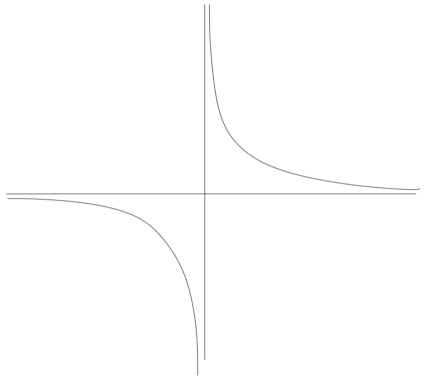


Figure 1.2: Projectively, the hyperbola is continuous as it crosses the $\vec{x}$ and $\vec{y}$ axes

Exercise 1.1 *: Consider the parabola $y = x^2$. Translate this into homogeneous coordinates and show that the line at infinity is tangent to it. Interpret the tangent geometrically by considering the parabola as the limit as k tends to ∞ of the ellipse $2kx^2 + (y - k)^2 - k^2 = 0$ (hint: this has tangent $y = 2k$).*

Exercise 1.2 *: Show that translation of a planar point by (a, b) is equivalent to multiplying its homogeneous coordinate column vector by*

$$\begin{pmatrix} 1 & 0 & a \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix}$$

Exercise 1.3 *: Show that multiplying the affine (*i.e. *inhomogeneous) coordinates of a point by a 2×2 matrix A is equivalent to multiplying its homogeneous coordinates by*

$$\left(\begin{array}{c|c} A & \begin{matrix} 0 \\ 0 \end{matrix} \\ \hline 0\ 0 & 1 \end{array}\right)$$

What is the homogeneous transformation matrix for a point that is rotated by angle θ about the origin, then translated by (a, b)?

1.2 The Perspective Camera

1.2.1 Perspective Projection

Following Dürer and the Renaissance painters, *perspective projection* can be defined as follows (see fig. 1.3). The *center of projection* is at the origin O of the 3D reference frame of the space. The image plane Π is parallel to the $(\vec{x}, \vec{y})$ plane and displaced a distance f (*focal length*) along the $\vec{z}$ axis from the origin. The 3D point P projects to the image point p. The orthogonal projection of O onto Π is the *principal point* o, and the $\vec{z}$ axis which corresponds to this projection line is the *principal axis* (sometimes called the *optical axis* by computer vision people, although there is no optic here at all).

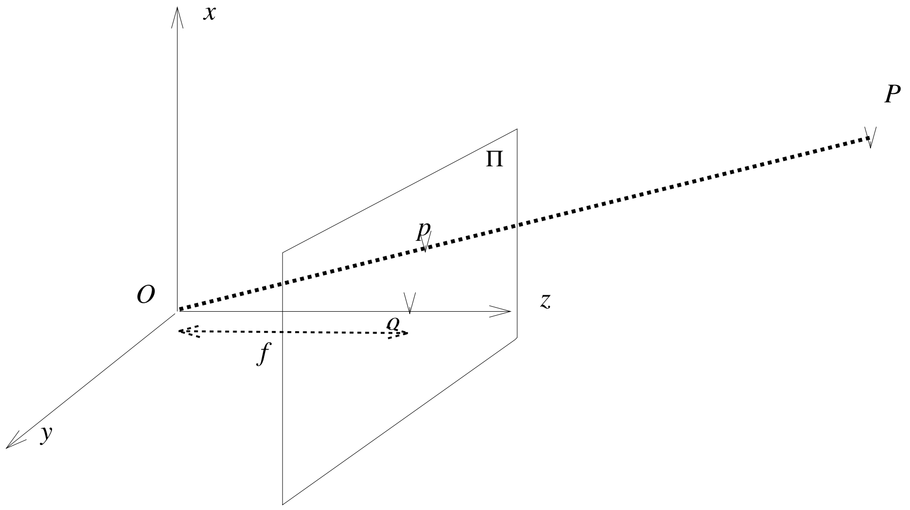


Figure 1.3: Standard perspective projection

Let (x, y) be the 2D coordinates of p and (X, Y, Z) the 3D coordinates of P. A direct application of Thales theorem shows that:

$$x = \frac{fX}{Z} \qquad y = \frac{fY}{Z}$$

We can assume that $f = 1$ as different values of f just correspond to different scalings of the image. Below, we will incorporate a full camera calibration into the model. In homogeneous coordinates, the above equations become:

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \sim \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

In real images, the origin of the image coordinates is not the principal point and the scaling along each image axis is different, so the image coordinates undergo a further transformation described by some matrix K. Also, the world coordinate system does not usually coincide with the perspective reference frame, so the 3D coordinates undergo a Euclidean motion described by some matrix M (see exercise 1.3), and finally we have:

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \sim K \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} M \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \tag{1.1}$$

M gives the 3D position and pose of the camera and therefore has six degrees of freedom which represent the *exterior* (or *extrinsic*) camera parameters. In a minimal parametrization, M has the standard 6 degrees of freedom of a rigid motion. K is independent of the camera position. It contains the *interior* (or *intrinsic*) parameters of the camera. It is usually represented as an upper triangular matrix:

$$K = \begin{pmatrix} s_x & s_\theta & u_0 \\ 0 & s_y & v_0 \\ 0 & 0 & 1 \end{pmatrix} \tag{1.2}$$

where s_x and s_y stand for the scalings along the $\vec{x}$ and $\vec{y}$ axes of the image plane, s_θ gives the skew (non-orthogonality) between the axes (usually $s_\theta \approx 0$), and (u_0, v_0) are the coordinates of the *principal point* (the intersection of the principal axis and the image plane).

Note that in homogeneous coordinates, the perspective projection model is described by linear equations: an extremely useful property for a mathematical model.

1.2.2 Real Cameras

Perspective (*i.e.* pinhole) projection is an idealized mathematical model of the behaviour of real cameras. How good is this model? — There are two aspects to this: extrinsic and intrinsic.

Light entering a camera has to pass through a complex lens system. However, lenses are designed to mimic point-like elements (pinholes), and in any case the camera and lens is usually negligibly small compared to the viewed region. Hence, in most practical situations the camera is "effectively point-like" and rather accurately satisfies the *extrinsic perspective* assumptions: *(i)* for each pixel, the set of 3D points projecting to the pixel (*i.e.* whose possibly-blurred images are centered on the pixel) is a straight line in 3D space; and *(ii)* all of the lines meet at a single 3D point (the *optical center*).

On the other hand, practical lens systems are nonlinear and can easily introduce significant distortions in the *intrinsic perspective* mapping from external optical rays to internal pixel coordinates. This sort of distortion can be corrected by a nonlinear deformation of the image-plane coordinates.

There are several ways to do this. One method, well known in the photogrammetry and vision communities, is to explicitly model the radial and decentering distortion (see [24]): if the center of the image is (u_0, v_0), the new coordinates (x', y') of the corrected point are given by

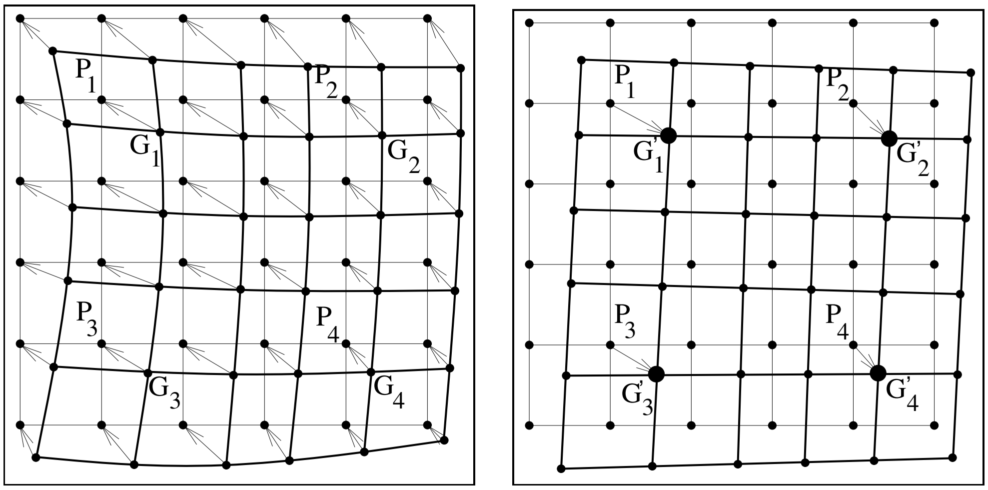
$$\begin{aligned} x' &= x + k_1\overline{x}r^2 + k_2\overline{x}r^4 + k_3\overline{x}r^6 + P_1(2\overline{x}^2 + r^2) + 2P_2\overline{xy} \\ y' &= y + k_1\overline{y}r^2 + k_2\overline{y}r^4 + k_3\overline{y}r^6 + P_2(2\overline{y}^2 + r^2) + 2P_1\overline{xy} \\ & \quad \text{where } \overline{x} = x - u_0, \;\; \overline{y} = y - u_0, \;\; r = \overline{x}^2 + \overline{y}^2 \end{aligned}$$

This linearizes the image geometry to an accuracy that can reach $2 \cdot 10^{-5}$ of the image size [1]. The first order radial distortion correction k_1 usually accounts for about 90% of the total distortion.

A more general method that does not require knowledge of the principal point and makes no assumptions about the symmetry of the distortion is based on a fundamental result in projective geometry:

Theorem: *In real projective geometry, a mapping is projective if and only if it maps lines onto either lines or points.*

Hence, to correct for distortion, all we need to do is to observe straight lines in the world and deform the image to make their images straight. Experiments described in [2] show accuracies of up to $1 \cdot 10^{-4}$ of the image for standard off-the-shelf CCD cameras. Figure 1.4 illustrates the process: line intersections are accurately detected in the image, four of them are selected to define a projective basis for the plane, and the others are re-expressed in this frame and perturbed so that they are accurately aligned. The resulting distortion corrections are then interpolated across the whole image. Careful



(a) Points are first located with respect to a four point projective basis

(b) A projective mapping then brings the points back to their aligned positions

Figure 1.4: Projective correction of distortion

experiments showed that an off-the-shelf (512×512) CCD camera with a standard frame-grabber could be stably rectified to a fully projective camera model to an accuracy of $1/20$ of a pixel.

Chapter 2

Basic Properties of Projective Space

2.1 Projective Space

Given a coordinate system, n-dimensional real *affine space* is the set of points parameterized by the set of all n-component real column vectors $(x_1, \ldots, x_n)^\top \in \mathbb{R}^n$.

Similarly, the points of real n-dimensional *projective space* $\mathbb{P}^n$ can be represented by $n+1$-component real column vectors $(x_1, \ldots, x_{n+1})^\top \in \mathbb{R}^{n+1}$, with the provisos that at least one coordinate must be non-zero and that the vectors $(x_1, \ldots, x_{n+1})^\top$ and $(\lambda x_1, \ldots, \lambda x_{n+1})^\top$ represent the same point of $\mathbb{P}^n$ for all $\lambda \neq 0$. The x_i are called *homogeneous coordinates* for the projective point.

2.1.1 Canonical Injection of $\mathbb{R}^n$ into $\mathbb{P}^n$

Affine space $\mathbb{R}^n$ can be embedded isomorphically in $\mathbb{P}^n$ by the standard injection $(x_1, \ldots, x_n) \longmapsto (x_1, \ldots, x_n, 1)$. Affine points can be recovered from projective ones with $x_{n+1} \neq 0$ by the mapping

$$(x_1, \ldots, x_{n+1}) \sim (\frac{x_1}{x_{n+1}}, \ldots, \frac{x_n}{x_{n+1}}, 1) \longmapsto (\frac{x_1}{x_{n+1}}, \ldots, \frac{x_n}{x_{n+1}})$$

A projective point with $x_{n+1} = 0$ corresponds to an ideal "point at infinity" in the $(x_1, ..., x_n)$ direction in affine space. The set of all such "infinite" points satisfying the homogeneous linear constraint $x_{n+1} = 0$ behaves like a hyperplane, called the hyperplane at infinity.

However, these mappings and definitions are affine rather than projective concepts. They are only meaningful if we are told in advance that $(x_1, \ldots, x_n)$ represents "normal" affine space and x_{n+1} is a special homogenizing coordinate. In a general projective space any coordinate (or linear combination) can act as the homogenizing coordinate and all hyperplanes are equivalent — none is especially singled out as the "hyperplane at infinity". These issues will be discussed more fully in chapter 4.

2.1.2 Projective Mappings

Definition: *A* nonsingular projective mapping *between two projective spaces is any mapping defined by multiplication of homogeneous coordinates by a full rank matrix. A* collineation *on $\mathbb{P}^n$ is an invertible projective mapping of $\mathbb{P}^n$ onto itself.*

All projective mappings can be represented by matrices. As with homogeneous coordinate vectors, these are only defined up to a non-zero rescaling.

Example: $I\!P^1 \longmapsto I\!P^1$. The general case of a collineation is:

$$\begin{pmatrix} x \\ t \end{pmatrix} \longmapsto \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax+by \\ cx+dy \end{pmatrix}$$

with $ad - cd \neq 0$. Provided $t \neq 0$ and $cx + d \neq 0$, this can be rewritten in inhomogeneous affine coordinates as:

$$\begin{pmatrix} x \\ 1 \end{pmatrix} \longmapsto \begin{pmatrix} ax+b \\ cx+d \end{pmatrix} \sim \begin{pmatrix} \frac{ax+b}{cx+d} \\ 1 \end{pmatrix}$$

Property: *A translation in affine space corresponds to a collineation leaving each point at infinity invariant.*

Proof: The translation $(x_1, ..., x_n, 1) \longmapsto (x_1 + a_1, ..., x_n + a_n, 1)$ can be represented by the matrix:

$$A = \left(\begin{array}{ccc|c} 1 & \cdots & 0 & a_1 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & a_n \\ \hline 0 & \cdots & 0 & 1 \end{array}\right)$$

Obviously $A(x_1, ..., x_n, 0)^\top = (x_1, ..., x_n, 0)$. □

More generally, any affine transformation is a collineation, because it can be decomposed into a linear mapping and a translation:

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} t_1 \\ \vdots \\ t_n \end{pmatrix}$$

In homogeneous coordinates, this becomes:

$$\left(\begin{array}{c} y_1 \\ \vdots \\ y_n \\ \hline 1 \end{array}\right) = \left(\begin{array}{c|c} & t_1 \\ A & \vdots \\ & t_n \\ \hline 0 \;\ldots\; 0 & 1 \end{array}\right) \left(\begin{array}{c} x_1 \\ \vdots \\ x_n \\ \hline 1 \end{array}\right)$$

Exercise 2.1 *: Prove that a collineation is an affine transformation if and only if it maps the hyperplane at infinity $x_{n+1} = 0$ into itself (*i.e. *all points at infinity are mapped onto points at infinity).*

Camera calibration: Assuming that the camera performs a exact perspective projection (see 1.1.2), we have seen that the image formation process can be expressed as a projective mapping from $I\!P^3$ to $I\!P^2$. *Projective camera calibration* is the computation of the projection matrix associated with this mapping. This is usually done using a set of points whose 3D locations $(X, Y, Z, T)^\top$ are known. If a pint projects to pixel coordinates (u, v), the projection equations can be written:

$$\begin{pmatrix} \lambda u \\ \lambda v \\ \lambda \end{pmatrix} = P \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

Taking ratios to eliminate the unknown scale factor λ, we have:

$$\begin{aligned} u &= \frac{p_{11}x + p_{12}y + p_{13}z + p_{14}}{p_{31}x + p_{32}y + p_{33}z + p_{34}} \\ v &= \frac{p_{21}x + p_{22}y + p_{23}z + p_{24}}{p_{31}x + p_{32}y + p_{33}z + p_{34}} \end{aligned} \tag{2.1}$$

As P is only defined up to an overall scale factor, this system has 11 unknowns. At least 6 points are required for a unique solution, but usually many more points are used in a least squares optimization that minimizes the effects of measurement uncertainty.

The projection matrix P contains both interior and exterior camera parameters. We will not consider the decomposition process here, as the exterior orientation/interior calibration distinction is only meaningful when projective space is reduced to Euclidean.

Exercise 2.2 *: Assuming perspective projection centered at the origin onto plane $z = 1$, and a $\vec{x}, \vec{y}$ image reference frame corresponding to the $\vec{x}, \vec{y}$ directions of the 3D reference frame, prove that the projection matrix P has the form*

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

The null space of P (the set of X such that $PX = 0$) corresponds to which 3D point X? What does the 3D point $(x, y, 0)$ project to?

2.1.3 Projective Bases

A *projective basis* for $\mathbb{P}^n$ is any set of $n+2$ points of $\mathbb{P}^n$, no $n+1$ of which lie in a hyperplane. Equivalently, the $(n+1) \times (n+1)$ matrix formed by the column vectors of any $n+1$ of the points must have full rank.

It is easily checked that $\{(1, 0, \ldots, 0)^\top, (0, 1, 0, \ldots, 0)^\top, \ldots, (0, \ldots, 0, 1)^\top, (1, \ldots, 1)^\top\}$ forms a basis, called the *canonical basis*. It contains the points at infinity along each of the n coordinate axes, the origin, and the *unit point* $(1, \ldots, 1)^\top$. Any basis can be mapped into this standard form by a suitable collineation.

Property: *A collineation on $\mathbb{P}^n$ is defined entirely by its action on the points of a basis.*

A full proof can be found in [23]. We will just check that there are the right number of constraints to uniquely characterize the collineation. This is described by an $(n+1) \times (n+1)$ matrix A, defined up to an overall scale factor, so it has $(n+1)^2 - 1 = n(n+2)$ degrees of freedom. Each of the $n+2$ basis point images $Ab_i \sim b'_i$ provides n constraints ($n+1$ linear equations defined up a common scale factor), so the required total of $n(n+2)$ constraints is met.

Exercise 2.3 *: Consider three non-aligned points a_i in the plane, and their barycenter g. Check that in homogeneous coordinates $(x, y, 1)$, we have*

$$g \sim \sum_{i=1}^{3} a_i$$

An analogous relation holds for the unit point in the canonical basis.

2.1.4 Hyperplanes and Duality

The Duality Principle

The set of all points in $\mathbb{R}^n$ whose coordinates satisfy a linear equation

$$a_1 X_1 + \ldots + a_n X_n + a_{n+1} = 0 \qquad \vec{X} \in \mathbb{R}^n$$

is called a *hyperplane*. Substituting homogeneous coordinates $X_i = x_i / x_{n+1}$ and multiplying out, we get a homogeneous linear equation that represents a *hyperplane in $\mathbb{P}^n$*:

$$(a_1, \ldots, a_{n+1}) \cdot (x_1, \ldots, x_{n+1}) = \sum_{i=1}^{n+1} a_i x_i = 0 \qquad \vec{x} \in \mathbb{P}^n \tag{2.2}$$

Notice the symmetry of equation (2.2) between the hyperplane coefficients $(a_1, ..., a_{n+1})$ and the point coefficients $(x_1, ..., x_{n+1})$. For fixed $\vec{x}$ and variable $\vec{a}$, (2.2) can also be viewed as the equation characterizing the hyperplanes $\vec{a}$ passing through a given point $\vec{x}$. In fact, the hyperplane coefficients $\vec{a}$ are also only defined up to an overall scale factor, so the space of all hyperplanes can be considered to be another projective space called the *dual* of the original space $\mathbb{P}^n$. By the symmetry of (2.2), the dual of the dual is the original space.

An extremely important *duality principle* follows from this symmetry:

Duality Principle: *For any projective result established using points and hyperplanes, a symmetrical result holds in which the roles of hyperplanes and points are interchanged: points become planes, the points in a plane become the planes through a point,* etc.

For example, in the projective plane, any two distinct points define a line (*i.e.* a hyperplane in 2D). Dually, any two distinct lines define a point (their intersection). Note that duality only holds universally in projective spaces: for example in the affine plane parallel lines do not intersect at all.

Desargues Theorem

Projective geometry was invented by the French mathematician Desargues (1591–1661) (for a biography in French, see *http://bib1.ulb.ac.be/coursmath/bio/desargue.htm*). One of his theorems is considered to be a cornerstone of the formalism. It states that "Two triangles are in perspective from a point if and only if they are in perspective from a line" (see fig. 2.1):

Theorem: *Let A, B, C and A', B, C' be two triangles in the (projective) plane. The lines AA', BB', CC' intersect in a single point if and only if the intersections of corresponding sides $(AB, A'B')$, $(BC, B'C')$, $(CA, C'A')$ lie on a single line.*

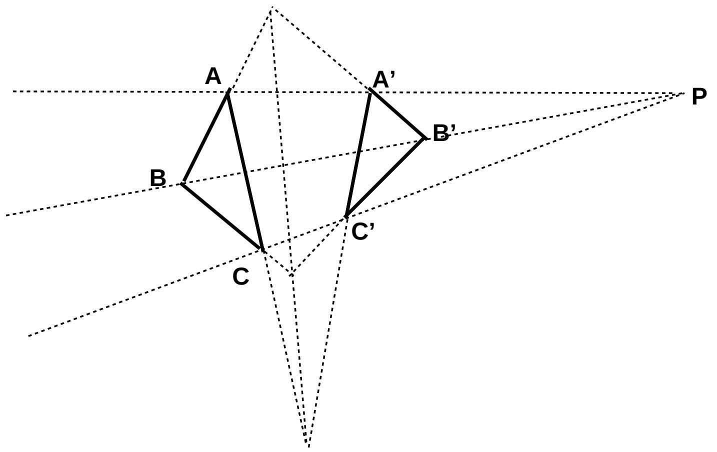


Figure 2.1: Two triangles in a Desargueian configuration

The theorem has a clear self duality: given two triplets of lines $\{a, b, c\}$ and $\{a', b', c'\}$ defining two triangles, the intersections of the corresponding sides lie on a line if and only if the lines of intersection of the corresponding vertices intersect in a point.

We will give an algebraic proof: Let P be the common intersection of AA', BB', CC'. Hence there are scalars $\alpha, \beta, \gamma, \alpha', \beta', \gamma'$ such that:

$$\left.\begin{array}{l} \alpha A - \alpha' A' = P \\ \beta B - \beta' B' = P \\ \gamma C - \gamma' C' = P \end{array}\right\} \Longrightarrow \left\{\begin{array}{l} \alpha A - \beta B = \alpha' A' - \beta' B' \\ \beta B - \gamma C = \beta' B' - \gamma' C' \\ \gamma C - \alpha A = \gamma' C' - \alpha' A' \end{array}\right.$$

This indicates that the point $\alpha A - \beta B$ on the line AB also lies at $\alpha' A' - \beta' B'$ on the line $A'B'$, and hence corresponds to the intersection of AB and $A'B'$, and similarly for $\beta B - \gamma C = BC \cap B'C'$ and $\gamma C - \alpha A = CA \cap C'A'$. But given that

$$(\alpha A - \beta B) + (\beta B - \gamma C) + (\gamma C - \alpha A) = 0$$

the three intersection points are linearly dependent, *i.e.* collinear. $\square$

Exercise 2.4 *: The sun (viewed as a point light source) casts on the planar ground the shadow $A'B'C'$ of a triangular roof ABC (see fig. 2.2). Consider a perspective image of all this, and show that it is a Desargueian configuration. To which 3D line does the line of intersections in Desargues theorem correspond? If a further point D in the plane ABC produces a shadow D', show that it is possible to reconstruct the image of D from that of D'.*

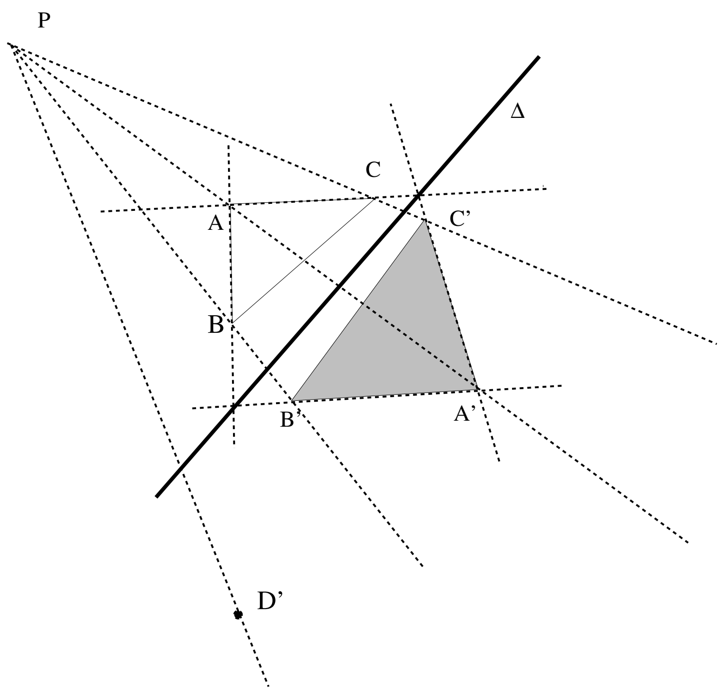


Figure 2.2: Shadow of a triangle on a planar ground

Hyperplane Transformations

In a projective space, a collineation can be defined by its $(n+1) \times (n+1)$ matrix M with respect to some fixed basis. If X and X' are the coordinate vectors of the original and transformed points, we have

$$X' = MX.$$

This maps hyperplanes of points to transformed hyperplanes, and we would like to express this as transformation of dual (hyperplane) coordinates. Let A and A' be the original and transformed hyperplane coordinates. For all points X we have:

$$AX = 0 \Longleftrightarrow A'X' = 0 \Longleftrightarrow A'MX = 0$$

The correct transformation is therefore $A'M = A$ or:

$$A' = AM^{-1}$$

or if we choose to represent hyperplanes by column vectors:

$$A'^t = (M^{-1})^t A^t$$

Of course, all of this is only defined up to a scaling factor. The matrix $M^* = (M^{-1})^t$ is sometimes called the *dual* of M.

2.2 Linear Algebra and Homogeneous Coordinates

This section gives some examples of the power of linear algebra on homogeneous coordinates as a tool for projective calculations.

2.2.1 Lines in the Plane and Incidence

We will develop the theory of lines in the projective plane $\mathbb{P}^2$. Most of the results can also be extended to higher dimensions.

Let $M = (x, y, t)^\top$ and $N = (u, v, w)^\top$ be homogeneous representatives for two distinct points in the plane (*i.e.* M and N are linearly independent 3-vectors: $M \neq \lambda N$). Let $L = (a, b, c)$ be a the dual coordinates of a line (hyperplane) in the plane. By the definition of a hyperplane, M lies on L if and only if the dot product $L \cdot M$ vanishes, *i.e.* if and only if the 3-vector L is orthogonal to the 3-vector M. The line MN through M and N must be represented by a 3-vector L orthogonal to both M and N, and hence proportional to the cross product $M \times N$:

$$L \sim M \times N = \begin{pmatrix} yw - tv \\ tu - xw \\ xv - yu \end{pmatrix}$$

Since $\times$ is bilinear, the mapping $M \to MN$ for fixed N is a linear mapping defined by the matrix

$$\begin{pmatrix} 0 & w & -v \\ -w & 0 & u \\ v & -u & 0 \end{pmatrix}$$

The vector N generates the kernel of this mapping.

Another way to characterize the line MN is as the set of points $P = \lambda M + \mu N$ for arbitrary λ, μ. Evidently

$$\begin{aligned} (M \times N) \cdot P &= \lambda (M \times N) \cdot M + \mu (M \times N) \cdot N \\ &= 0 + 0 = 0 \end{aligned}$$

Dually, if L and L' are two lines defined by their dual coordinates, then $\lambda L + \mu L'$ is some other line through the intersection X of L and L' (since $L \cdot X = 0 = L' \cdot X$ implies $(\lambda L + \mu L') \cdot X = 0$). As $\lambda : \mu$ varies, this line traces out the entire pencil of lines through X. By duality, $X = L \times L'$.

One further way to approach lines is to recall that if M, N and P are collinear, each point is a linear combination of the two others. In particular, the 3×3 determinant $|MNP|$ vanishes. If M and N are fixed, this provides us with a linear constraint that P must satisfy if it is to lie on the line: $|MNP| = (M \times N) \cdot P = 0$.

2.2.2 The Fixed Points of a Collineation

A point A is fixed by a collineation with matrix H exactly when $HA \sim A$, *i.e.* $HA = \alpha A$ for some scalar α. In other words, A must be a right eigenvector of H. Since an $(n+1) \times (n+1)$ matrix typically has $n+1$ distinct eigenvalues, a collineation in $\mathbb{P}^n$ typically has $n+1$ fixed points, although some of these may be complex.

H maps the line through any two fixed points A and B onto itself: $H(\lambda A + \mu B) = \alpha\lambda A + \beta\mu B$. In addition, if A and B have the same eigenvalue ($\alpha = \beta$), $\lambda A + \mu B$ is also an eigenvector and the entire line AB is pointwise invariant. In fact, the pointwise fixed subspaces of $\mathbb{P}^n$ under H correspond exactly to the eigenspaces of H's repeated eigenvalues (if any).

Exercise 2.5 : *Show that the matrix associated with a plane translation has a triple eigenvalue, but that the corresponding eigenspace is only two dimensional. Provide a geometric interpretation of this.*

Exercise 2.6 : *In 3D Euclidean space, consider a rotation by angle θ about the $\vec{z}$ axis. Find the eigenvalues and eigenspaces, prove algebraically that the rotation axis is pointwise invariant, and show that in addition the "circular points" with complex coordinates $(1, i, 0, 0)^\top$ and $(1, -i, 0, 0)^\top$ are fixed.*

Chapter 3

Projective Invariants & the Cross Ratio

Following Klein's idea of studying geometry by its transformations and invariants, this chapter focuses on the invariants of projective geometry. These form the basis of many other results and help to provide intuition about the structure of projective space. The simplest projective invariant is the *cross ratio*, which generates a scalar from four points of any 1D projective space (*e.g.* a projective line).

3.1 Some Standard Cross Ratios

3.1.1 Cross-Ratios on the Projective Line

Let M and N be two distinct points of a projective space. The dimension of the underlying space is irrelevant: they might be points in the projective line, plane or 3D space, hyperplanes, *etc*. The *projective line between M and N* consists of all points A of the form

$$A = \lambda M + \mu N$$

Here (λ, μ) are the coordinates of A in the 2D linear subspace spanned by the coordinate vectors M and N. Projectively, (λ, μ) are only defined up to an overall scale factor, so they really represent homogeneous coordinates on the abstract projective line $\mathbb{P}^1$ from M to N, expressed with respect to the linear basis of coordinate vectors $\{M, N\}$.

Let $A_i, i = 1, \ldots, 4$ be any four points on this line. Their cross ratio $\{A_1, A_2; A_3, A_4\}$ is defined to be:

$$\begin{aligned} \{A_1, A_2; A_3, A_4\} &= \frac{(\lambda_1\mu_3 - \lambda_3\mu_1)(\lambda_2\mu_4 - \lambda_4\mu_2)}{(\lambda_1\mu_4 - \lambda_4\mu_1)(\lambda_2\mu_3 - \lambda_3\mu_2)} \\ &= \frac{(\lambda_1/\mu_1 - \lambda_3/\mu_3)(\lambda_2/\mu_2 - \lambda_4/\mu_4)}{(\lambda_1/\mu_1 - \lambda_4/\mu_4)(\lambda_2/\mu_2 - \lambda_3/\mu_3)} \end{aligned} \tag{3.1}$$

$a/0 = \infty$ is a permissable value for the cross ratio. If the numerator and denominator both vanish at least three of the λ/μ must be identical, so by l'Hôpital's rule the cross ratio is 1. The key property of the cross ratio is that it is invariant under collineations and changes of basis. In other words, it is a *projective invariant*.

Theorem: *The cross ratio does not depend on the choice of basis M and N on the line. If H is a collineation, then $\{A_1, A_2; A_3, A_4\} = \{HA_1, HA_2; HA_3, HA_4\}$.*

See [23] for detailed proofs. In fact, invariance under collineations need only be verified on the projective line $\mathbb{P}^1$:

Exercise 3.1 *: On the projective line, collineations are represented by nonsingular* 2×2 *matrices:*

$$\begin{pmatrix} \lambda' \\ \mu' \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \lambda \\ \mu \end{pmatrix} \qquad ad - cd \neq 0$$

Show explicitly that the cross ratio is collineation invariant (Hint: $\lambda\mu' - \lambda'\mu = det\binom{\lambda\ \lambda'}{\mu\ \mu'}$*).*

Cross Ratios and Length Ratios: In (λ, μ) coordinates on the line MN, N is represented by $(0, 1)$ and serves as the "origin" and M is represented by $(1, 0)$ and serves as the "point at infinity". For an arbitrary point $A = (\lambda, \mu) \not\sim (1, 0)$, we can rescale (λ, μ) to $\mu = 1$, and represent A by its "affine coordinates"$(\lambda, 1)$, or just λ for short. Since we have mapped M to infinity, this is just linear distance along the line from N. Hence, setting $\mu_i = 1$ in 3.1, the cross ratio becomes a ratio of length ratios. The ancient Greek mathematicians already used cross ratios in this form.

Exercise 3.2 *: Let D be the point at infinity on the projective line, and let A, B, C be three finite points. Show that*

$$\{A, B; C, D\} = \frac{\overline{AC}}{\overline{BC}}$$

Exercise 3.3 *If MN is the line between two points M and N in the projective plane, use the (λ, μ) parameterization and the results of section 2.2.1 to show that the cross ratio of four points A, B, C, D on the line is*

$$\{A, B; C, D\} = \frac{(A \times C) \cdot (B \times D)}{(A \times D) \cdot (B \times C)}$$

Cross Ratios and Projective Bases: In section 2.1 we saw that any 3 distinct points on the projective line can be used as a projective basis for it. (NB: if we also specify their homogeneous scales, only two are needed, as with M and N above). The cross ratio $k = \{A, B; C, D\}$ of any fourth point D with the points of a projective basis A, B, C defines D uniquely with respect to the basis. Rescaling to $\mu_i = 1$ as above, we have

$$\lambda_4 = \frac{(\lambda_1 - \lambda_3)\lambda_2 + (\lambda_3 - \lambda_2)\lambda_1 k}{(\lambda_1 - \lambda_3) + (\lambda_3 - \lambda_2)k} \tag{3.2}$$

As this is invariant under projective transformations, the cross ratio can be used to invariantly position a point with respect to a projective basis on a line. A direct application is reconstruction of points on a 3D line using measured image cross ratios.

Exercise 3.4 *: In an image, we see three equally spaced trees in a line. Explain how the position of a fourth tree in the sequence can be predicted. If the first tree lies at $\lambda = 0$ on the image line, the second lies at 1 and the third at c, what is the image coordinate of the fourth tree?*

3.1.2 Cross Ratios of Pencils of Lines

Let U and V be two lines in the projective plane, defined by their dual coordinate vectors. Consider a set of lines $\{W_i\}$ defined by:

$$W_i = \lambda_i U + \mu_i V$$

with the (λ_i, μ_i) defined up to scale as usual. $\{W_i\}$ belongs to the *pencil of lines* through the intersection of U and V (*c.f.* section 2.2.1), so each W_i passes through this point. As in section 3.1.1, the cross ratio $\{W_1, W_2; W_3, W_4\}$ is defined to be the cross ratio of the four homogeneous coordinate pairs (λ_i, μ_i). Dually to points, we have:

Theorem: *The cross ratio of any four lines of a pencil is invariant under collineations.*

Cross ratios of collinear points and coincident lines are linked as follows:

Theorem: *The cross ratio of four lines of a pencil equals the cross ratio of their points of intersection with an arbitrary fifth line transversal to the pencil (*i.e. *not through the pencil's centre) — see fig. 3.1.*

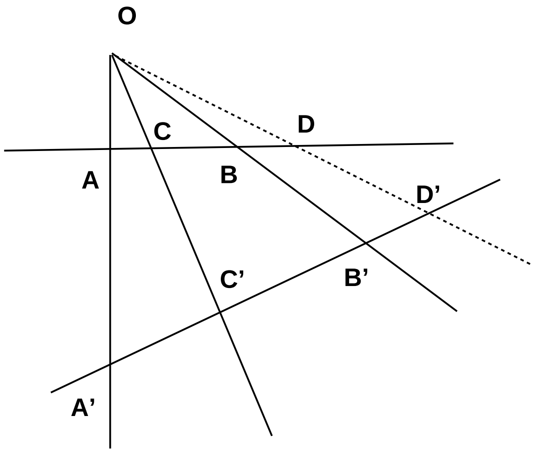


Figure 3.1: The cross ratio of four lines of a pencil

In fact, we already know that the cross ratios of the intersection points must be the same for any two transversal lines, since the lines correspond bijectively to one another under a central projection, which is a collineation.

The simplest way to establish the result is to recall the line intersection formulae of section 2.2.1. If U and V are the basis 3-vectors for the line pencil, and L is the transversal line, the intersection points of L with U, V and $\lambda U + \mu V$ are respectively $L \times U$, $L \times V$ and $\lambda(L \times U) + \mu(L \times V)$. In other words, the (λ, μ) coordinates of a line in the basis $\{U, V\}$ are the same as the (λ, μ) coordinates of its intersection with L in the basis $\{L \times U, L \times V\}$. Hence, the two cross ratios are the same.

Another way to prove the result is to show that cross ratios of lengths along the transversal line can be replaced by cross ratios of angle sines, and hence are independent of the transversal line chosen (*c.f* fig. 3.1):

$$\frac{\overline{AC} \cdot \overline{BC}}{\overline{AD} \cdot \overline{BD}} = \frac{\sin(OA, OC)\ \sin(OB, OD)}{\sin(OA, OD)\ \sin(OB, OC)} \tag{3.3}$$

However, this is quite a painful way to compute a cross ratio. A more elegant method uses determinants:

Theorem (Möbius): *Let $L_i, i = 1, \ldots, 4$ be any four lines intersecting in O, and $A_i, i = 1, \ldots, 4$ be any four points respectively on these lines, then*

$$\{L_1, L_2; L_3, L_4\} = \frac{\mid OA_1A_3 \mid\ \mid OA_2A_4 \mid}{\mid OA_1A_4 \mid\ \mid OA_2A_3 \mid} \tag{3.4}$$

where $\mid OA_iA_j \mid$ denotes the determinant of the 3×3 matrix whose columns are the homogeneous coordinate vectors of points O, A_i and A_j.

This 19th century result extends gracefully to higher dimensions. To prove it, let $(a, b, 1)$, $(x, y, 1)$ and $(u, v, 1)$ be the normalized affine coordinate vectors of O, A_i and A_j. Then

$$|OA_iA_j| = \begin{vmatrix} a & x & u \\ b & y & v \\ 1 & 1 & 1 \end{vmatrix} = \begin{vmatrix} a & x-a & u-a \\ b & y-b & v-b \\ 1 & 0 & 0 \end{vmatrix}$$

$$= \quad \vec{OA_i} \times \vec{OA_j} = |OA_i|\ |OA_j| \cdot \sin(OA_i, OA_j)$$

The vector's lengths cancel out of the cross ratio of these terms, and if the coordinate vectors do not have affine normalization, the scale differences cancel out too. $\square$

Projective Bases for the Projective Plane: Any four distinct coplanar points (no three of which are collinear) form a projective basis for the plane they span. Given four such points A, B, C, D, a fifth point M in the plane can be characterized as the intersection of one line of the pencil through D with one line of the pencil through B (see fig. 3.2). Hence, M can be parameterized by two cross ratios (one for each pencil). This construction fails when M lies on the line DB: in this case, another family of pencils has to be considered. This is a common phenomenon in projective spaces: a single system of coordinates does not cover the entire space without singularities or omissions.

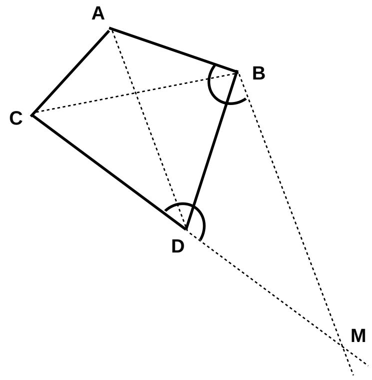


Figure 3.2: Locating a fifth point in a plane

Four-point bases can be used directly for visual reconstruction of coplanar 3D points. Given four known points on a 3D plane and their perspective images, a fifth unknown point on the plane can be reconstructed from its image by expressing the image in terms of the four known image points, and then using the same homogeneous planar coordinates (λ, μ, ν) to re-create the corresponding 3D point.

3.1.3 Cross Ratios of Planes

A *pencil of planes* in $\mathbb{P}^3$ is a family of planes having a common line of intersection. The cross ratio of four planes Π_i of a pencil is the same as the cross ratio of the lines l_i of intersection of the planes with fifth, transversal plane (see fig. 3.3). Once again, different transversal planes give the same cross ratio as the figures they give are projectively equivalent. The Möbius formula also extends to this case: let P, Q be any two distinct points on the axis of the plane pencil, and A_i, $i = 1, \ldots, 4$ be points lying on each plane Π_i (not on the axis), then

$$\{\Pi_1, \Pi_2; \Pi_3, \Pi_4\} = \frac{|\ PQA_1A_3\ |\ \ |\ PQA_2A_4\ |}{|\ PQA_1A_4\ |\ \ |\ PQA_2A_3\ |}$$

where $|\ PQA_iA_j\ |$ stands for a 4×4 determinant of 4-component column vectors.

3.2 Harmonic Ratios and Involutions

3.2.1 Definition

Let A, B, C, D be four points on a line with cross ratio k. From the definition of the cross ratio, it follows that the $4! = 24$ possible permutations of the points yield 6 different (but functionally

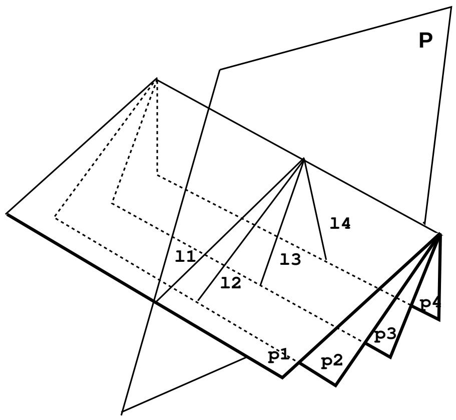


Figure 3.3: A pencil of planes

equivalent) cross ratios:

$$\begin{aligned}\{A, B; C, D\} &= \{B, A; D, C\} \\ &= \{C, D; A, B\} \\ &= \{D, C; B, A\} \\ &= 1/\{A, B; D, C\} = k \\ \{A, C; B, D\} &= 1/\{A, C; D, B\} = 1 - k \\ \{A, D; B, C\} &= 1/\{A, D; C, B\} = 1 - 1/k\end{aligned}$$

In some symmetrical cases, the six values reduce to three or even two. When two points coincide, the possible values are 0, 1 and ∞. With certain symmetrical configurations of complex points the possible values are $-e^{2\pi i/3}$ and $-e^{-2\pi i/3}$. Finally, when $k = -1$ the possible values are -1, $1/2$ and 2. The Ancient Greek mathematicians called this a *harmonic configuration*. The couples (A, B) and (C, D) are said to be *harmonic pairs*, and we have the following equality:

$$\{A, B; C, D\} = \{A, B; D, C\} = \{B, A; C, D\} = \{C, D; A, B\} = -1.$$

Hence, the harmonic cross ratio is invariant under interchange of points in each couple and also (as usual) under interchange of couples, so the couples can be considered to be unordered. Given (A, B), D is said to be *conjugate* to C if $\{A, B; C, D\}$ forms a harmonic configuration.

Exercise 3.5 *: If C is at infinity, where is its harmonic conjugate?*

The previous exercise explains why the harmonic conjugate is considered to be a projective extension of the notion of an affine mid-point.

3.2.2 The Complete Quadrangle

Any set of four non-aligned points A, B, C, D in the plane can be joined pairwise by six distinct lines. This figure is called the *complete quadrangle*. The intersections of opposite sides (including the two diagonals) produce three further points E, F, G (see fig. 3.4).

Property: *The pencil of four lines based at each intersection of opposite sides is harmonic. For example* $\{FA, FB; FE, FG\} = -1$.

A simple proof maps the quadrangle projectively onto a rectangle, sending F and G to points at infinity in two orthogonal directions. The four lines of the pencil become parallel, with one being at infinity. The line through E is obviously half way between those through A and B, so by the previous exercise the line configuration is harmonic. □

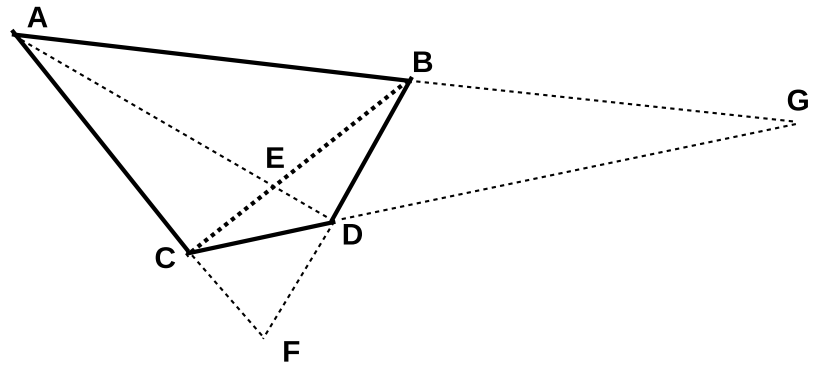


Figure 3.4: The complete quadangle

Exercise 3.6 *: Given the images a, b, c of two points A, B on a 3D line and its 3D vanishing point (point at infinity) C, construct the projection of the mid-point of AB using just a ruler.*

3.2.3 Involutions

Definition: *An* involution *is non-trivial projective collineation H whose square is the identity:* $H^2 = \mathrm{Id}$.

Property: *The mapping taking a point to its harmonic conjugate is an involution.*

By the definition of a harmonic pair, the mapping is its own inverse. Using the cross ratio formula (3.2), it is straightforward to check that it is also projective. □

Exercise 3.7 *: Consider the affine line,* i.e. *the projective line with the point at infinity fixed. Show that the only affine involutions (projective involutions fixing the point at infinity) are reflections about a fixed point (*e.g. *$x \to -x$, reflection about 0). Deduce that on the projective line, all involutions with one fixed point have two fixed points, and thence that they map points to their harmonic conjugates with respect to these two fixed points.*

Exercise 3.8 *: (NB: This is an algebraic continuation of the previous exercise). Let a collineation H be represented by a 2×2 matrix as in exercise 3.1. Prove that if H is an involution then $a = -d$. Show that H always has two distinct fixed points, either both real or both complex. Give an H that fixes the complex points $(i, 1)^\top$ and $(-i, 1)^\top$, and show that if $x' = Hx$, then $\{A, B; x, x'\} = -1$. Given the invariance of the cross ratio under complex collineations, show that all involutions of the projective line map points to their harmonic conjugates with respect to the two fixed points of the involution.*

Involutions are useful when dealing with images of symmetrical objects. Consider the simple case of a planar object with reflection symmetry, as in fig. 3.5. In 3D, the lines joining corresponding points of the object are parallel. In the image, all such lines meet at the projection of the associated point at infinity S. The projection of the line of symmetry consists of points conjugate to S with respect of pairs of corresponding points on the shape. There is a plane involution of the image mapping points on the object to their opposites, and fixing S and each point of the symmetry axis. Also, any line through any two points of the figure meets the line through the two opposite points on the axis: hence the axis is easy to find given a few correspondences. All this extends to 3D symmetries: see [22] chapter 8.

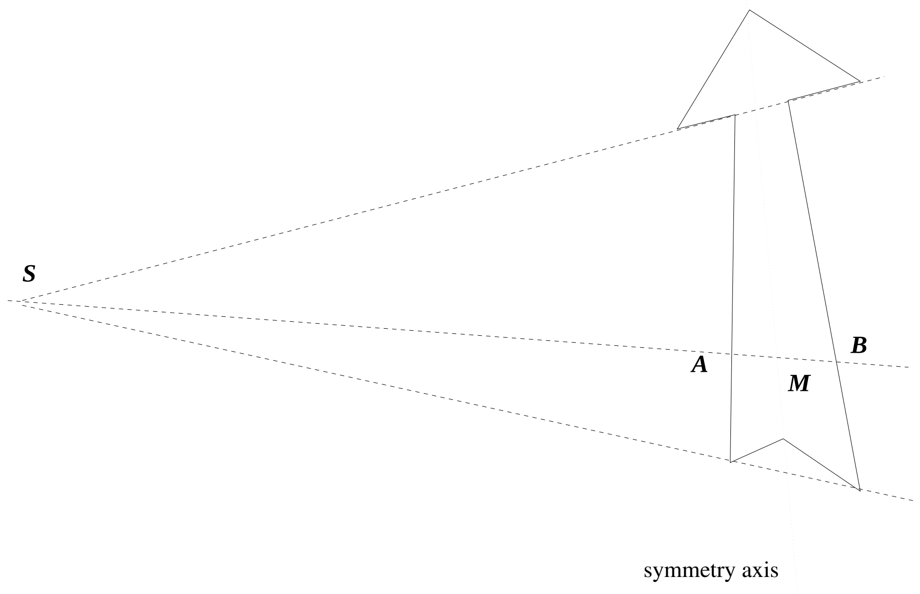


Figure 3.5: Perspective view of a planar symmetrical shape

3.3 Recognition with Invariants

Invariant measures like the cross ratio contain shape information. This can be used in several ways, for instance to identify objects seen by a camera. For simplicity, we will restrict ourselves to planar objects here, so that the mapping from scene points to image ones is one to one. The full 3D case is harder. For example, no 3D invariant can be extracted from a single perspective image of a set of general 3D points, unless something further is known about the point configuration (for instance, that some of them are coplanar). For 3D invariants from two images, see [7]. For an overview of projective invariants for planar shape recognition, see [22].

3.3.1 Generalities

Four aligned points provide a projective invariant, the cross ratio. If we only allow affine transformations, three aligned points are enough to provide an invariant: the ratio of their separations. Finally, limiting ourselves to rigid planar motions, there is an invariant for only two points: their separation. In other words, invariants are only defined relative to a *group* $\mathcal{G}$ of permissable transformations. The smaller the group, the more invariants there are. Here we will mainly concentrate on the group of planar projective transformations.

Invariants measure the properties of configurations that remain constant under arbitrary group transformations. A configuration may contain many invariants. For instance 4 points lead to 6 different cross ratios. However only one of these is *functionally independent*: once we know one we can trivially calculate the other five. In general, it is useful to restrict attention to functionally independant invariants.

An important theorem tells us how many independant invariants exist for a given algebraic configuration. Let $\mathcal{C}$ be the configuration and $\mathrm{dof}(\mathcal{C})$ its number of degrees of freedom, *i.e.* the number of independant parameters needed to describe it, or the dimension of the corresponding parameter manifold. Similarly, let $\mathrm{dof}(\mathcal{G})$ be the number of independent parameters required to characterize a transformation of group $\mathcal{G}$; let $\mathcal{G}_\mathcal{C}$ be the subgroup of $\mathcal{G}$ that leaves the configuration invariant as a shape; and let $\mathrm{dof}(\mathcal{G}_\mathcal{C})$ be this *isotropy subgroup*'s number of degrees of freedom. Then we have the following theorem [6]:

Theorem: *The number of independant invariants of a configuration $\mathcal{C}$ under transformations $\mathcal{G}$ is*

$$dof(\mathcal{C}) - dof(\mathcal{G}) + dof(\mathcal{G}_{\mathcal{C}})$$

Consider a few examples:

1. Let $\mathcal{G}$ be the 2 dof group of affine transformations on the line. No continuous family of affine transformations leaves 3 collinear points (which have 3 dof) invariant: $\text{dof}(\mathcal{G}_{\mathcal{C}}) = 0$. So there is $3 - 2 + 0 = 1$ independant invariant, *e.g.* one of the length ratios.

2. Let $\mathcal{G}$ be the 3 dof group of rigid transformations in the plane and $\mathcal{C}$ be the 3 dof configuration of two parallel lines. $\mathcal{G}_{\mathcal{C}}$ is the 1 dof subgroup of translations parallel to this line, so there is $3 - 3 + 1 = 1$ independant invariant, *e.g.* the distance between the lines.

3. Two conics in the projective plane have $2 \cdot 5 = 10$ dof. The planar projective group has 8 dof and no continuous family of projective transformations leaves two conics globally invariant; therefore there are $10 - 8 + 0 = 2$ invariants for such a configuration.

In practice, the isotropy subgroup $(\mathcal{G}_{\mathcal{C}})$ often reduces to the identity. The assumption it has 0 dof is known as the "counting argument". However, it can be very difficult to spot isotropies, so care is needed:

Exercise 3.9 *: In the projective plane, consider two lines L, L' and two points A, B in general position. How many invariants does the counting argument suggest? However there is at least one invariant: A, B and the intersections of AB with L and L' define a cross ratio. Exhibit the isotropy subgroup and show that it has 1 dof. (Hint: Map A, B to infinity and the intersection of the lines to the origin: what is the isotropy subgroup now?)*

3.3.2 Five Coplanar Points

Now consider the simple case of 5 coplanar points in general position. This example will show that although the theory provides a nice framework, it does not give all the answers.

The isotropy subgroup is the identity: even four of the points are sufficient to define a unique projective basis. There are therefore $5 \cdot 2 - 8 + 0 = 2$ independent projective invariants. In section 3.1.2, we showed that two suitable invariants can be obtained by taking cross ratios of pencils of lines through any two of the points.

Ideally, we would like to be able to recognise one among a set of such configurations by computing the two cross ratios from image data and searching a database for the closest known configuration with those invariant values. This raises two questions:
1) Which of the $5! = 120$ possible orderings of the image points was used for the invariant stored in the database?
2) How should proximity of cross ratios be measured?

The first point is a combinatorial/correspondence problem. One way to attack this is to create combinations of invariants that are unchanged under permutations of the points. For example, the 6 possible values of a cross ratio can be combined into a symmetric, order independent form such as

$$k^2 + \frac{1}{k^2} + (1-k)^2 + \frac{1}{(1-k)^2} + (1 - \frac{1}{k})^2 + (\frac{k}{k-1})^2$$

Exercise 3.10 *: The simplest symmetrical polynomials would be the sum or product of all the values. Why are these not useful?*

Even given this, we still have to choose two of the five points as base points for the symmetrized cross ratios. Again we can form some symmetric function of the $5 \cdot 4 = 20$ possiblities. We could take the maximum or minimum, or some symmetric polynomial such as:

$$I_1 = \sum_{i=1}^{5} a_i, \qquad I_2 = \sum_{i \neq j} a_i a_j$$

The problem with this is that each time we symmetrize the invariants lose discriminating power.

To compare invariants, something like the traditional Mahalanobis distance can be used. However, most projective invariants are highly nonlinear, so the distance has to be evaluated separately at each configuration: using a single overall distance threshold ϵ usually gives very bad results. In fact, close to degenerate configurations — here, when three of the points are almost aligned — even well-designed Mahalanobis metrics may be insufficient. Such configurations have to be processed case-by-case (usually, they are discarded).

Hence, even with careful design, the results of this indexation approach are not very satisfactory. An exhaustive test on randomly generated point configurations was performed in [17], with the conclusion that no usable trade off between false positives and false negatives existed.

However, the results can be improved by several order of magnitude by the following process:

1. Given that convexity and order is preserved under perspective image projections (although not under general projective mappings!), classify the 5 point configurations into one of the three classes described in fig. 3.6

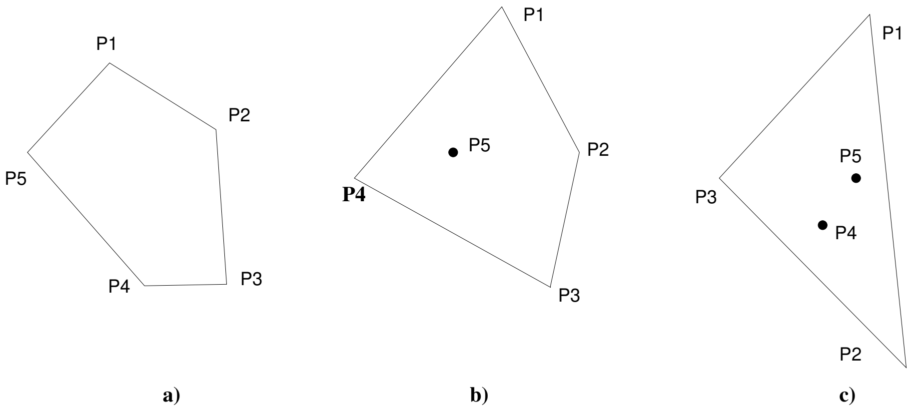


Figure 3.6: Three topologically different configurations for 5 points

2. For each class, compute the possible cross ratios. For instance, for class (a) there are five possibilities for the five vertices of the polygon, with the points considered in clockwise order.

3. For each set of (redundant) invariants, compute the Mahalanobis distance, index, and perform final classification by projective alignment with each of the retrieved candidate configurations.

Ordering the points allows the use of raw (non-symmetrized) cross ratios, which significantly improves the discriminating power. It also makes the invariant computation specific to each class of object considered. With random noise of 1.5 pixel in a 512×512 image, this process prunes on average about 799 of every 800 configurations. These results have also been validated by independent experiments [14].

Chapter 4

A Hierarchy of Geometries

We saw in chapter 2 that there is a standard mapping from the usual Euclidean space into projective space. The question is, if only the projective space is known, what additional information is required to get back to Euclidean space. We will only consider the case of the plane and 3D space, but the extension to arbitrary dimensions is straightforward. For more detailed expositions in the same spirit as this tutorial, see [5] for a geometric presentation and [15] for a computer vision oriented one.

In his Erlanger Programme (1872), Felix Klein formulated geometry as the study of a space of points together with a group of mappings, the geometric transformations that leave the structure of the space unchanged. Theorems are then just invariant properties under this group of transformations. Euclidean geometry is defined by the group of rigid displacements; similarity or extended Euclidean geometry by the group of similarity transforms (rigid motions and uniform scalings); affine geometry by the the group of affine transforms (arbitrary nonsingular linear mappings plus translations); and projective geometry by projective collineations.

There is a clear hierarchy to these four groups:

$$\text{Projective} \supset \text{Affine} \supset \text{Similarity} \supset \text{Euclidean}$$

As we go down the hierarchy, the transformation groups become smaller and less general, and the corresponding spatial structures become more rigid and have more invariants. We consider each step of the hierarchy in the following sections, specializing each geometry to the next in turn.

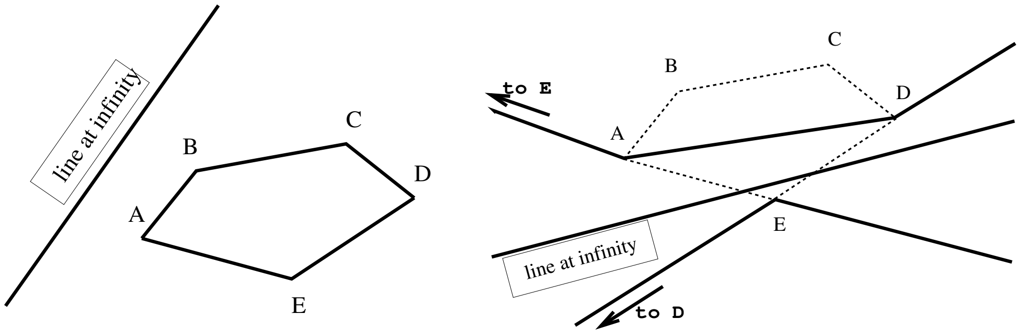
4.1 From Projective to Affine Space

4.1.1 The Need for Affine Space

Projective geometry allows us to discuss coplanarity, and relative position using the cross ratio or its derivatives. However in standard projective space there is no consistent notion of *betweenness*. For instance, we can not uniquely define the line segment linking two points A, B. The problem is that projective lines are topologically circular: they close on themselves when they pass through infinity (except that infinity is not actually distinguished in projective space — all points of the line are equal). So there are two equally valid segments linking two points on a projective line: $A + \lambda B$ for λ in either $[0, +\infty]$ and in $[-\infty, 0]$.

One solution to this problem is to distinguish a set (in fact a hyperplane) of points at infinity in projective space: this gives us affine space.

A related problem occurs if we try to define the convex hull of a set of points. Figure 4.1 illustrates the behavior of the convex hull for two different affine interpretations of the plane, *i.e.* for two different choices of the line at infinity. In the first case, the line at infinity (the ideal line) does not pass through the intuitive convex hull (fig. 4.1a), so the convex hull remains similar to the intuitive one. But in figure 4.1b the line cuts the intuitive convex hull, so B and C turn out to be inside the calculated hull in this case. (No side of the calculated convex hull ever cuts the line at infinity). Note that in this case, we need only know whether the ideal line passes through the desired convex hull or not; this naturally extends to the ideal plane in the 3D space (see [21]).



a) b)

Figure 4.1: The convex hull depends on the position of the hyperplane at infinity

Another important limitation when working with projective geometry is that it does not allow us to define the midpoint of a segment: the midpoint is defined by a simple ratio and projective geometry deals only with cross ratios.

4.1.2 Defining an Affine Restriction

As we have seen in chapter 2, affine transformations leave the line at infinity invariant. In fact the group of affine transformations is just the subgroup of the projective group that maps this line (or plane) onto itself. Any line can be chosen to be fixed, and then the projective subgroup that leaves this line fixed defines a valid affine space.

Usually we want to interpret the fixed line as being the effective line at infinity of a hidden affine space. For example, this happens when a 3D plane is projected to an image under perspective projection. The line at infinity of the 3D plane is projected to a corresponding "horizon line", which is not usually at infinity in the image. Once this horizon line has been identified and fixed, the affine geometry defined is the one pertaining to the original 3D plane, and not to the camera image plane. To implement this, we just need to find a change of coordinates that maps this line to the one with dual coordinate vector $(0, 0, 1)$.

Sometimes (as with the real horizon) this horizon line is actually visible in the image. If not, we might use some prior knowledge about the scene to compute its position. For instance, if we know that two 3D lines are parallel, their intersection directly gives the image of one point at infinity. Another useful piece of information is the 3D distance ratio of three aligned points. The cross ratio of the ideal point on the line through these is the same as the distance ratio in affine space, so the image of the ideal point can be computed from (3.2), given the affine ratio and the projections of the three points.

Exercise 4.1 : *You observe the perspective image of a plane parallelogram. Derive an algebraic expression for the plane's horizon (*i.e. *the projection of the line at infinity) in terms of the vertex projections a, b, c, d.*

Exercise 4.2 : *You observe the image of three parallel, equally spaced, coplanar lines. Explain how to compute the horizon line of the plane.*

Once the ideal line (or plane) has been located in projective space, computing affine information is straightforward. Stating that Π_∞ should have equation $t = 0$ yields 2 independent linear constraints on the last line of the permissible collineation matrices W (3 in 3D).

4.2 From Affine to Euclidean Space

Here, by Euclidean space we actually mean space under the group of similarity transforms, *i.e.* we allow uniform changes of scale in addition to rigid displacements. This permits a very elegant algebraic formulation, and in any case scale can never be recovered from images, it can only be derived from prior knowledge or calibration. (You can never tell from images whether you are looking at the real world or a reduced model). In practice, Euclidean information is highly desirable as it allows us to measure angles and length ratios.

Last century, Laguerre showed that Euclidean structure is given by the location in the plane at infinity Π_∞ of a distinguished conic, whose equation in a Euclidean coordinate system is

$$\begin{cases} x^2 + y^2 + z^2 = 0 \\ t = 0 \end{cases} \tag{4.1}$$

This is known as the *absolute conic* Ω [13, 23]. All points lying on it have complex coordinates so it is a little difficult to picture, but for the most part it behaves just like any other conic.

Exercise 4.3 : *Show that the absolute conic Ω is mapped onto itself under scaled Euclidean transformations. From there, show that the corresponding image conic is invariant under rigid displacements of the camera, provided that the camera's internal parameters remain unchanged.*

As in the projective to affine case, prior Euclidean information is needed to recover Euclidean structure from affine space. Perhaps the easiest way to do this is to reconstruct known circles in 3D space. Algebraically, each such circle intersects Π_∞ in exactly two complex points, and these always belong to Ω [23]. Ω itself can be reconstructed from three such circles. Let the resulting equation be

$$X^t Q X = (x, y, z) \begin{pmatrix} a_1 & a_4 & a_5 \\ a_4 & a_2 & a_6 \\ a_5 & a_6 & a_3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0$$

A change of coordinates is needed to bring Ω into the form of equation (4.1). As the matrix Q is symmetric, there is an orthogonal matrix P such that:

$$Q = P^t \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} P$$

Setting $X' = PX$, we have:

$$X^t Q X = (X')^t \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} X'$$

With a further rescaling along each axis, we get equation (4.1).

Another way to proceed is to use the basic extended Euclidean invariant: the angle α between two coplanar lines L and L'. Such angles also put a constraint on Ω and can be used to compute it. Let A

and A' be the intersections of the two lines with Π_∞. Let I and J be the intersections with Ω of the (line at infinity in the) plane defined by these two lines. Laguerre's formula states that:

$$\alpha = \frac{1}{2i} \log(\{A, A'; I, J\})$$

We can write $I = A + tA'$, $J = A + t'A'$. With this notation, $\{A, A'; I, J\} = t/t' = e^{2i\alpha}$. If we require that both I and J lie on Ω we get the constraint equations

$$\begin{aligned} t^2 A'^T Q A' + 2t A^T Q A' + A^T Q A &= 0 \\ e^{4i\alpha} t^2 A'^T Q A' + 2e^{2i\alpha} t A^T Q A' + A^T Q A &= 0 \end{aligned} \tag{4.2}$$

A polynomial constraint on Q is easily derived. Eliminate t^2 between the above equations:

$$2t A^T Q A' \beta(\beta - 1) + A^T Q A(\beta^2 - 1) = 0 \quad \text{with} \quad \beta = e^{2i\alpha}$$

extract t from this

$$t = -\frac{(\beta+1)}{2\beta} \frac{A^T Q A}{A^T Q A'}$$

and substitute into (4.2) to provide a quadratic polynomial constraint on Q:

$$(1 + \beta)(A^T Q A)(A'^T Q A') - 2(A^T Q A')^2 = 0$$

Theoretically, the absolute conic Ω can be computed from the above constraint given 5 known angles. However, in practice there does not seem to be a closed form solution and in our experiments we have used different Euclidean constraints. But the above discussion does clearly show the relationships between the different layers of projective, affine and Euclidean reconstruction, and specifies exactly what structure needs to be obtained in each case.

4.3 Summary

Projective space $\mathbb{P}^n$ is invariant under the $n(n+2)$ parameter *projective group* of $(n+1) \times (n+1)$ matrices up to scale (8 dof in 2D, 15 dof in 3D). The fundamental projective invariant is the cross ratio which requires four objects in a projective pencil (one parameter configuration, or homogeneous linear combination of two basis objects $\lambda X + \mu Y$). Projective space has notions of subspace incidence and an elegant duality between points and hyperplanes, but no notion of rigidity, distance, points 'at infinity', or sidedness/betweenness. It is the natural arena for perspective images and uncalibrated visual reconstruction.

N-dimensional *affine space* is invariant under the $n(n+1)$ parameter *affine group* of translations and linear deformations (rotations, non-isotropic scalings and skewings). Affine space is obtained from projective space by fixing an arbitrary hyperplane to serve as the hyperplane of points 'at infinity': requiring that this be fixed puts n constraints on the allowable projective transformations, reducing them to the affine subgroup. The fundamental affine invariant is ratio of lengths along a line. Given 3 aligned points A, B, C, the length ratio AB/AC is given by the cross ratio $A, B; D, C$, where D is the line's affine point at infinity. Affine space has notions of 'at infinity', sidedness/betweenness, and parallelism (lines meeting at infinity), but no notion of rigidity, angle or absolute length.

Similarity or *scaled Euclidean space* is invariant under the $n(n+1)/2 + 1$ parameter *similarity group* of rigid motions (rotations and translations) and uniform scalings. The fundamental similarity invariants are angles and arbitrary length ratios (including non-aligned configurations). Euclidean space is obtained from affine space by designating a conic in the hyperplane at infinity to serve as the *absolute conic*. The similarity group consists of affine transformations that leave the $n(n+1)/2 - 1$

parameters of this conic fixed. Angles can be expressed using the cross ratio and properties of this conic. Scaled Euclidean space has all the familiar properties of conventional 3D space, except that there is no notion of scale or absolute length.

Fixing this final scale leaves us with the $n(n+1)/2$ parameter *Euclidean group* of rigid motions of standard *Euclidean space*.

Chapter 5

Projective Stereo vision

The projective approach to stereo vision investigates what can be done with completely uncalibrated cameras. This is very important, not only because it frees us from the burden of calibration for certain tasks, but also because it provides mathematical background for many aspects of stereo vision and multiple image algebra, such as self calibration techniques.

The subject was perhaps investigated by mathematical photogrammetrists in the 19th century, but if so the results do not seem to have survived. More recently, following an unnoticed paper by Thompson [25] in 1968, work on projective invariants around 1990 and the 1992 projective reconstruction papers of Faugeras [3] and Hartley [9] launched an enormous burst of research in this field.

5.1 Epipolar Geometry

5.1.1 Basic Considerations

Consider the case of two perspective images of a rigid scene. The geometry of the configuration is depicted in fig. 5.1. The 3D point M projects to point m in the left image and m' in the right one. Let O and O' be the centres of projection of the left and right camera respectively. The M's *epipolar*

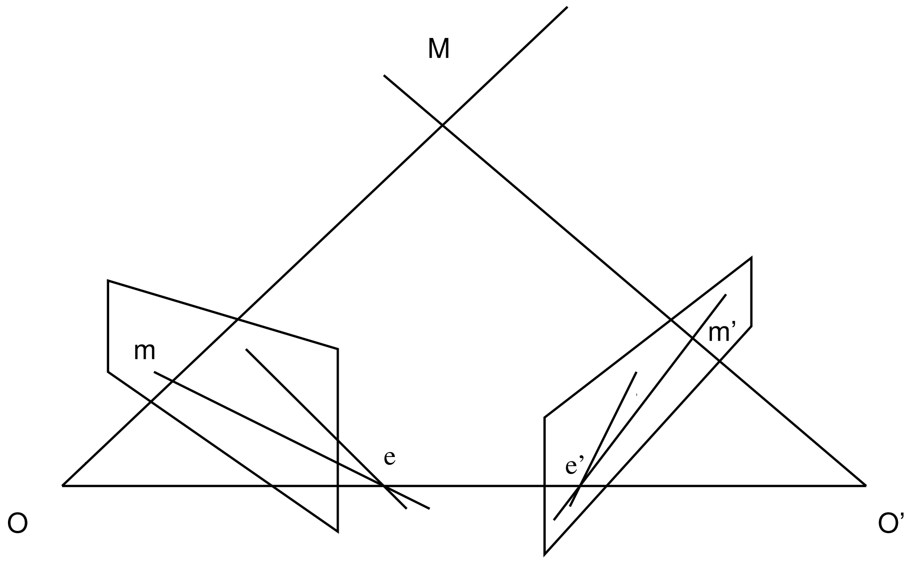


Figure 5.1: Epipolar geometry

plane $[O, O', M]$ intersects the image planes in two conjugate lines called M's *epipolar lines*. These pass respectively through the images m and m' and the points e and e' of intersection of the image planes with the base line $[O, O']$. These conjugate points are called the *epipoles* of the stereo rig.

Now let M move around in space. The epipolar planes form a pencil of planes through $[O, O']$, and the epipolar lines form two pencils of lines through e and e' respectively.

Key Property: *The epipolar line-line and line-plane correspondences are projective.*

Proof: Perspective projection through any third point on $[O, O']$ provides the required projective collineation. □

Projectively, the epipolar geometry is all there is to know about a stereo rig. It establishes the correspondence $m \leftrightarrow m'$, and allows 3D reconstruction of the scene to be carried out up to an overall 3D projective deformation (which is all that *can* be done with any number of completely uncalibrated cameras, without further constraints). An important practical application of epipolar geometry is to aid the search for corresponding points, reducing it from the entire second image to a single epipolar line. The epipolar geometry is sometimes obtained by calibrating each of the cameras with respect to the same 3D frame, but as the next section shows, it can easily be found from a few point correspondences, without previous camera calibration.

5.1.2 The Fundamental Matrix

Let $m = (x, y, t)^t$ be the homogeneous coordinates of a point in the first image and $e = (u, v, w)$ be the coordinates of the epipole of the second camera in the first image. The epipolar line through m and e is represented by the vector $l = (a, b, c)^t = m \times e$ (*c.f.* section 2.2.1). The mapping $m \to l$ is linear and can be represented by a 3×3 rank 2 matrix C:

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} yw - zv \\ zu - xw \\ xv - yu \end{pmatrix} = \begin{pmatrix} 0 & w & -z \\ -w & 0 & u \\ z & -u & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \tag{5.1}$$

The mapping of epipolar lines l from image 1 to the corresponding epipolar lines l' in image 2 is a collineation defined on the 1D pencil of lines through e in image 1. It can be represented (non-uniquely) as a collineation on the entire dual space of lines in $\mathbb{P}^2$. Let A be one such collineation: $l' = Al$.

The constraints on A are encapsulated by the correspondence of 3 distinct epipolar lines. The first two correspondences each provide two constraints, because a line in the plane has 2 dof. The third line must pass through the intersection of the first two, so only provides one further constraint. The correspondence of any further epipolar line is then determined, for example by its cross ratio with the three initial lines. Since A has eight degrees of freedom and we only have five constraints, it is not fully determined. Nevertheless, the matrix $F = AC$ *is* fully determined. Using (5.1) we get

$$l' = ACm = Fm \tag{5.2}$$

F is called the *fundamental matrix*. As C has rank 2 and A has rank 3, F has rank 2. The right kernel of C — and hence F — is obviously the epipole e. The fact that all epipolar lines in the second image pass through e' (*i.e.* ${e'}^t \cdot l' = 0$ for all transferred l') shows that the left kernel of F is e': ${e'}^t F = 0$.

F defines a bilinear constraint between the coordinates of corresponding image points. If m' is the point in the second image corresponding to m, it must lie on the epipolar line $l' = Fm$, and hence ${m'}^t \cdot l' = 0$ (*c.f.* 5.1). The epipolar constraint can therefore be written:

$${m'}^t F m = 0 \tag{5.3}$$

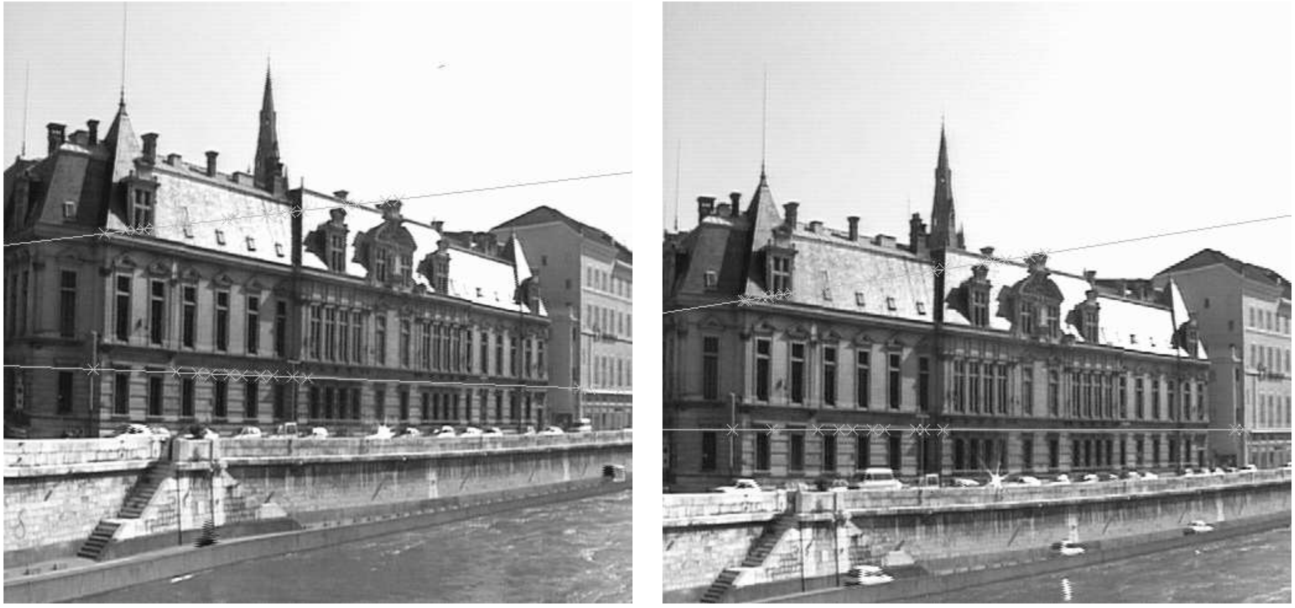


Figure 5.2: An example of automatic epipolar geometry computation

5.1.3 Estimating the Fundamental Matrix

To obtain the epipolar geometry, we need to estimate F. This can be done using some initial point correspondences between the images. Equation (5.3) shows that each matching pair of points between the two images provides a single linear constraint on F. This allows F to be estimated linearly (up to the usual arbitrary scale factor) from 8 independent correspondences. However, F as defined has only seven degrees of freedom: 2 from C (the epipole position) and 5 from A. Algebraically, F has 9 linear coefficients modulo one overall scale factor, but the rank 2 condition implies the additional constraint $\det(F) = 0$. Hence, F can actually be computed from only 7 matches in general position plus the rank constraint. However, since the latter is a cubic there will generally be three possible solutions in this case.

Figure 5.2 shows what can be done with a good automatic epipolar geometry algorithm. However, for good results, some care is needed. The discussion below is inspired by a paper of R. Hartley [8].

Assume that we have already found some matches $m \leftrightarrow m'$ between the images. Each correspondence provides a linear constraint on the coefficients of F: $m'^t F m = 0$. Expanding, we get:

$$xx'f_{11} + xy'f_{12} + xf_{13} + yx'f_{21} + yy'f_{22} + yf_{23} + x'f_{31} + y'f_{32} + f_{33} = 0$$

where the coordinates of m and m' are respectively $(x, y, 1)^t$ and $(x', y', 1)^t$. Combining the equations obtained for each match gives a linear system that can be written $Af = 0$, where f is a vector containing the 9 coefficients of F, and each row of A is built from the coordinates m and m' of a single match. Since F is defined only up to an overall scale factor, we can restrict the solution for f to have norm 1. We usually have more than the minimum number (8) of points, but these are perturbed by noise so we will look for a least squares solution:

$$\min_{||f||=1} ||Af||^2$$

As $||Af||^2 = f^t A^t A f$, this amounts to finding the eigenvector associated with the smallest eigenvalue of the 9×9 symmetric, positive semidefinite normal matrix $A^t A$. There are standard numerical techniques for this [20]. However, this formulation does not enforce the rank constraint, so a second step must be added to the computation to project the solution F onto the rank 2 subspace. This can be done by taking the Singular Value Decomposition of F and setting the smallest singular value to zero. Basically, SVD decomposes F in the form

$$F = QDR$$

where D is diagonal, and Q and R are orthogonal. Setting the smallest diagonal element of D to 0 and reconstituting gives the desired result.

The above method is standard, but if applied naïvely it is quite unstable. A typical image coordinate in a 512×512 image might be ~ 200. Some of the entries in a typical row of A are $xx' \sim 200^2$, others are $x \sim 200$, and the last entry is $1 \sim 1$, so there is a variation in size of $\sim 200^2$ among the entries of A, and hence of $200^4 \sim 2 \cdot 10^9$ among the entries of A^tA. This means that numerically, A^tA is extremely ill-conditioned: the solution contains an implicit least squares trade off, but it is nothing like the trade off we would actually like for maximum stability.

A simple solution to this is to normalize the pixel coordinates from $[0, 512]$ to $[-1, 1]$ before proceeding. This provides a well-balanced matrix A and much more stable and accurate results for F. In a practical implementation, a considerable effort must also be spent on rejecting false correspondences in the input data [27].

In summary, the procedure for estimating the epipolar geometry is:
— Extract points from the images
— Find an initial set of point correspondences (typically using correlation)
— Use the above fundamental matrix estimation algorithm
— Embed everything in a robust estimation framework resistant to outliers (*e.g* using Least Median Squares).

For a detailed discussion of alternatives to this scheme, see [12].

5.2 3D Reconstruction from Multiple Images

Suppose that a fixed scene is seen by two or more perspective cameras. We are interested in geometric issues, so we will suppose that the correspondences between visible points in different images are already known. However it must be pointed out that matching is a fundamental and extremely difficult issue in vision, which can not be dismissed so lightly in practice.

So in this section, we suppose that n 3D points A_i are observed by m cameras with projection matrices $P_j, j = 1, \ldots, m$. Neither the point positions nor the camera projections are given. Only the projections a_{ij} of the i^{th} point in the j^{th} image are known.

5.2.1 Projective Reconstruction

Simple parameter counting shows that we have $2nm$ independent measurements and only $11m + 3n$ unknowns, so with enough points and images the problem ought to be soluble. However, the solution can never be unique as we always have the freedom to change the 3D coordinate system we use. In fact, in homogeneous coordinates the equations become

$$a_{ij} \sim P_j A_i \qquad i = 1, \ldots n, \;\; j = 1, \ldots, m \tag{5.4}$$

So we always have the freedom to apply a nonsingular 4×4 transformation H to both projections $P_j \to P_j H^{-1}$ and world points $A_i \to HA_i$. Hence, without further constraints, reconstruction is only ever possible up to an unknown projective deformation of the 3D world. However, modulo this fundamental ambiguity, the solution *is* in general unique.

One simple way to obtain the solution is to work in a projective basis tied to the 3D points [3]. Five of the visible points (no four of them coplanar) can be selected for this purpose.

Exercise 5.1 *: Given the epipolar geometry, show how we can decide whether four points are coplanar or not, by just considering their images in a stereo pair. Hint: consider the intersection of a pair of image lines, each linking two of the four points.*

An alternative to this is to select the projection center of the first camera as the coordinate origin, the projection center of the second camera as the unit point, and complete the basis with three other visible 3D points A_1, A_2, A_3 such that no four of the five points are coplanar.

Exercise 5.2 *: Design an image-based test to check whether three points are coplanar with the center of projection. Derive a test that checks that two points are not coplanar with the base line of a stereo pair, assuming that the epipolar geometry is known. Deduce a straightforward test to check that the above five points form a valid 3D projective basis.*

Let a_1, a_2, a_3 and a'_1, a'_2, a'_3 respectively be the projections in image 1 and image 2 of the 3D points A_1, A_2, A_3. Make a projective transformation of each image so that these three points and the epipoles become a standard basis:

$$a_1 = a'_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \qquad a_2 = a'_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \qquad a_3 = a'_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \qquad e = e' = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Also fix the 3D coordinates of A_1, A_2, A_3 to be respectively

$$A_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \qquad A_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \qquad A_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$$

It follows that the two projection matrices can be written:

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \qquad P' = \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{pmatrix}$$

Exercise 5.3 *: Show that the projections have these forms. (NB: Given only the projections of A_1, A_2, A_3, each row of P, P' could have a different scale factor since point projections are only defined up to scale. It is the projections of the epipoles that fix these scale factors to be equal).*

Since the projection matrices are now known, 3D reconstruction is relatively straightforward. This is just a simple, tutorial example so we will not bother to work out the details. In any case, for precise results, a least squares fit has to be obtained starting from this initial algebraic solution (*e.g.* by bundle adjustment).

5.2.2 Affine Reconstruction

Section 4.1 described the advantages of recovering affine space and provided some methods of computing the location of the plane at infinity Π_∞. The easiest way to proceed is to use prior information, for instance the knowledge that lines in the scene are parallel or that a point is the half way between two others.

Prior constraints on the camera motion can also be used. For example, a translating camera is equivalent to a translating scene. Observing different images of the same point gives a line in the direction of motion. Intersecting several of these lines gives the point at infinity in the motion direction, and hence one constraint on the affine structure.

On the other hand, any line through two scene points translates into a line parallel to itself, and the intersection of these two lines gives further constraints on the affine structure. Given three such point pairs we have in all four points at infinity, and the projective reconstruction of these allows the ideal points to be recovered — see [16] for details and experimental results.

5.2.3 Euclidean Reconstruction

We have not implemented the specific suggestions of section 4.2 for the recovery of Euclidean structure from a projective reconstruction by observing known scene angles, circles,... However we have developed a more brute force approach, finding a projective transformation H in equation (5.4) that maps the projective reconstruction to one that satisfies a set of redundant Euclidean constraints. This simultaneously minimizes the error in the projection equations and the violation of the constraints. The equations are highly nonlinear and a good initial guess for the structure is needed. In our experiments this was obtained by assuming parallel projection, which is linear and allows easy reconstruction using SVD decomposition ([26]).

Figures 5.3 and 5.4 show an example of this process.

5.3 Self Calibration

This section is based on an unpublished paper written by Richard Hartley in 1995. It derives the Kruppa equations which allow a camera's internal parameters to be derived from matches between several views taken with the same camera at different positions and orientations. From there, Euclidean reconstruction of the scene up to an overall scale factor is possible without further prior information.

First, we present the geometric link between the absolute conic and camera's interior calibration. Then, we motivate and derive the Kruppa equations which provide two constraints on the calibration for each pair of views observed by the camera from different locations.

5.3.1 The Absolute Conic and the Camera Parameters

Consider two camera projections P and P' corresponding to the same camera (internal parameters) but different poses. We saw in exercise 4.3 that the image of the absolute conic (IAC) is independent of the camera pose. In fact, the IAC is directly related to the internal parameter matrix K of the camera defined in equation (1.2).

Given that the absolute conic is in any case invariant under rotations and translations, we can choose coordinates so that the first projection matrix reduces to $P = K(I_{3\times3}|0) = (K|0)$ (*c.f.* equation 1.1). Hence, for a point $X = (\vec{x}, 0)^t$ on the plane at infinity we have projection $u = K\vec{x}$ or $\vec{x} = K^{-1}u$. u is on the image of the absolute conic exactly when X is on the conic itself, *i.e.* when

$$u^t K^{-t} K^{-1} u = \vec{x}^t \vec{x} = 0$$

So the image of the absolute conic is given by the matrix $K^{-t}K^{-1}$. If this matrix can be found, by Choleski factorization we can extract K^{-1} and thence find the internal parameter matrix K. In fact, as we will see, it is easier to work from the outset with the inverse K^tK of the IAC matrix, called the *dual image of the absolute conic* (DIAC).

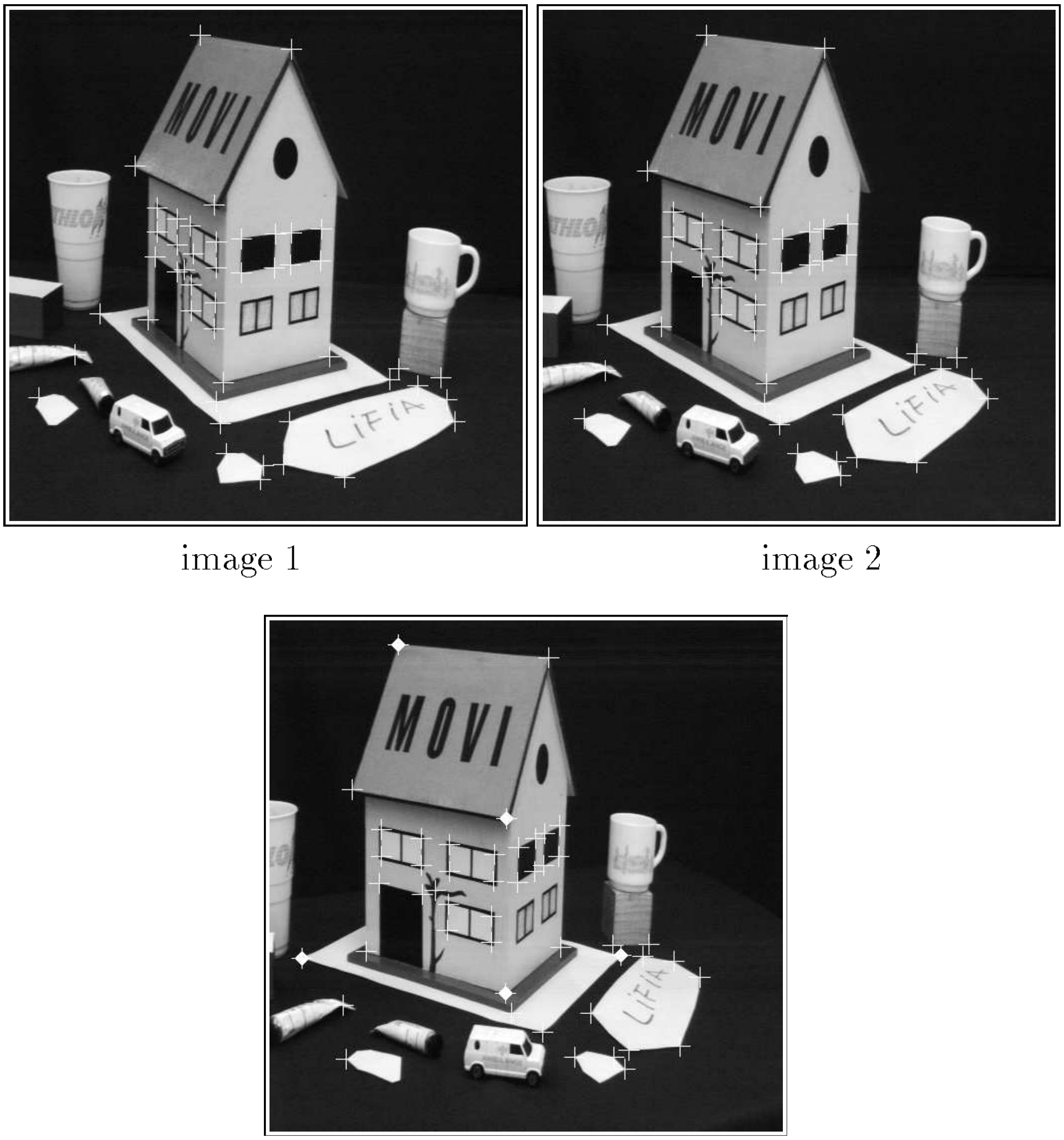


Figure 5.3: The house scene: the three images used for reconstruction together with the extracted corners. The reference frame used for the reconstruction is defined by the five points marked with white disks in image 3

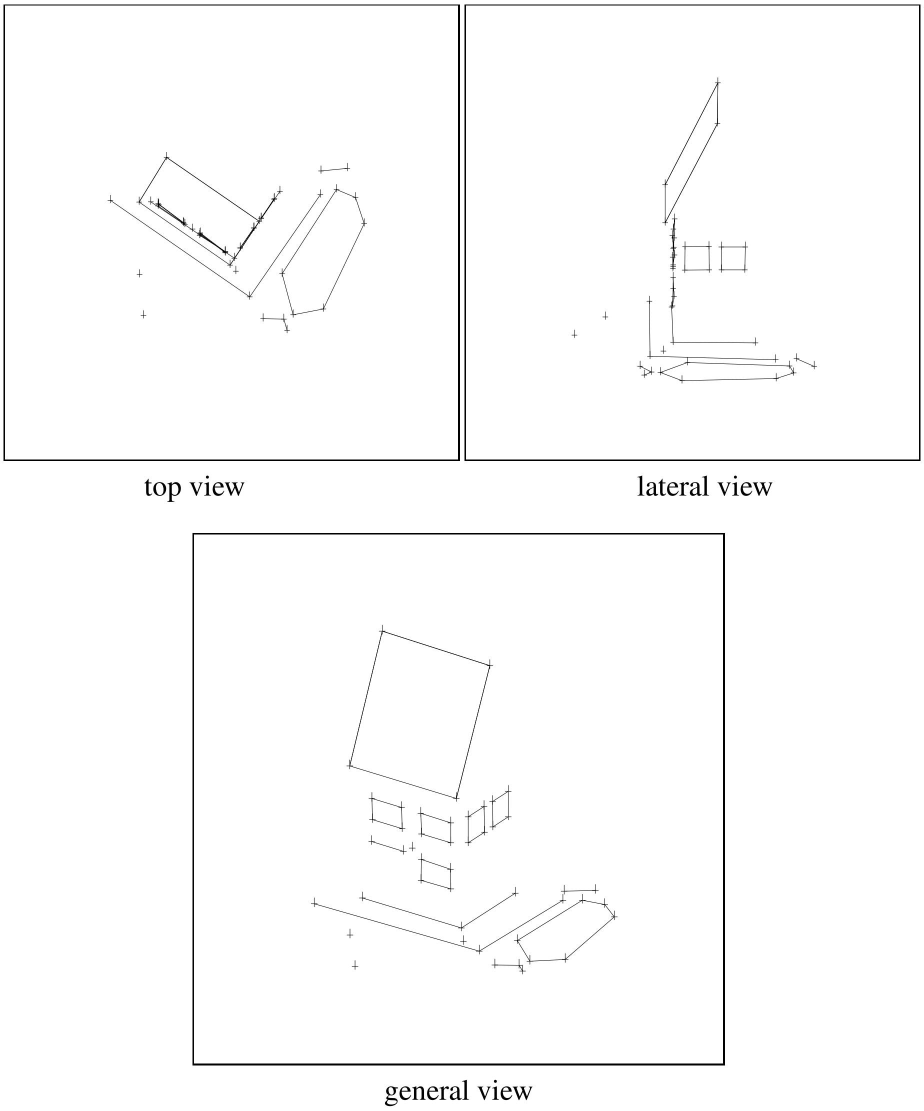


Figure 5.4: Euclidean reconstruction of an indoor scene using the known relative positions of five points. To make the results easier to see, the reconstructed points are joined with segments.

5.3.2 Derivation of Kruppa's Equations

Let F be the fundamental matrix of the stereo rig. Change image coordinates so that the epipoles are at the origins, and so that corresponding epipolar lines have identical coordinates. Hence, the last row and column of F vanish, and a short calculation shows that it has the form:

$$F' = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Let A and A' be the 3×3 matrices associated with this coordinate transformation in respectively the first and second image. K becomes respectively AK and $A'K$ and the projection matrices become AP and $A'P'$.

Consider a plane passing through the two camera centers and tangent to the absolute conic. This plane projects to a corresponding pair of epipolar lines in the images, which are tangent to the IAC's by construction. In fact there are two such tangential planes, so there are two pairs of corresponding epipolar tangents to the IAC's (see fig. 5.5).

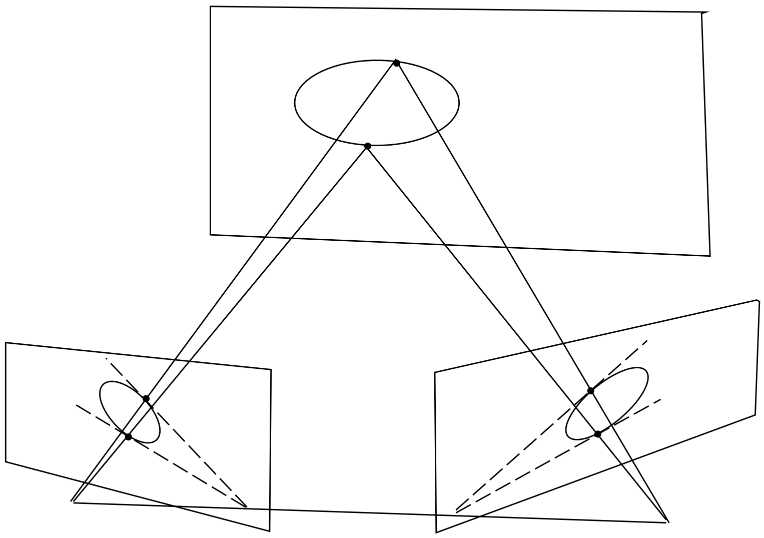


Figure 5.5: The two epipolar tangent planes to the absolute conic define two pairs of corresponding epipolar tangents to the IAC's

Tangent lines to conics are most easily expressed in terms of dual conics ([23]). The dual conic defines the locus of lines tangent to the original conic, expressed in terms of their dual (line 3-vector) coordinates. In a coordinate frame, the dual of a conic is given by the inverse of the original conic matrix. A line $(u, v, w)^t$ is tangent to a point conic defined by symmetric matrix C if and only if $(u, v, w)C^{-1}(u, v, w)^t = 0$, in other words, if and only if the line vector belongs to the dual conic C^{-1}.

Hence, we consider the DIAC's or duals of the images of the absolute conic $D = K^t A^t AK$ and $D' = K^t A'^t A'K$. After our change of coordinates, corresponding epipolar lines have identical coordinates and all pass through the origin. Let $l = (\lambda, \mu, 0)^t$ be an epipolar tangent to the IAC. Then in each image we have a tangency constraint: $l^t Dl = 0$ and $l^t D'l = 0$. Expanding and using the symmetry of D and D' gives

$$\begin{aligned} \lambda^2 d_{11} + 2\lambda\mu d_{12} + \mu^2 d_{22} &= 0 \\ \lambda^2 d'_{11} + 2\lambda\mu d'_{12} + \mu^2 d'_{22} &= 0 \end{aligned}$$

Since the tangents are corresponding epipolar lines, these two quadratics must have the same solutions for (λ, μ). Hence, they must be identical up to a scale factor:

$$\frac{d_{11}}{d'_{11}} = \frac{d_{12}}{d'_{12}} = \frac{d_{22}}{d'_{22}} \tag{5.5}$$

These are the Kruppa equations.

5.3.3 Explicit Computation

To derive explicit expressions for the matrices D and D' in terms of the fundamental matrix F, let us reconsider the above argument. Let $F = UWV^t$ be the Singular Value Decomposition of F. Here, U and V are orthogonal, and W is a diagonal matrix with diagonal values $r, s, 0$. We can write this as follows:

$$F = U \begin{pmatrix} r & & \\ & s & \\ & & 1 \end{pmatrix} \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} V^t$$

Define

$$A'^t = U \begin{pmatrix} r & & \\ & s & \\ & & 1 \end{pmatrix} \qquad A = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} V^t$$

Hence, $F = A'^t F' A$ with A, A' non-singular and F' having the desired canonical form. Applying the transformations $p \to Ap, p' \to A'p'$ and $F \to F' = A'^{-t} F A^{-1}$, we see that $p'Fp$ is unchanged and F becomes canonical, so A, A' are the required rectifying transformations. These transformations take $P \to AP$, $P' \to A'P'$ and hence $K \to AK$, $K \to A'K$, and hence the DIAC $C = KK^t$ becomes respectively $D = ACA^t$ and $D' = A'CA'^t$.

Now explicitly compute the d_{ij} in order to use equation (5.5). Decompose A and A' by rows:

$$A = \begin{pmatrix} \boldsymbol{a}_1^t \\ \boldsymbol{a}_2^t \\ \boldsymbol{a}_3^t \end{pmatrix} \qquad A' = \begin{pmatrix} \boldsymbol{a'}_1^t \\ \boldsymbol{a'}_2^t \\ \boldsymbol{a'}_3^t \end{pmatrix}$$

Then, $D = ACA^t$ implies $d_{ij} = \boldsymbol{a}_i^t C \boldsymbol{a}_j$, and we have the following explicit form for the Kruppa equations:

$$\frac{\boldsymbol{a}_1^t C \boldsymbol{a}_1}{\boldsymbol{a'}_1^t C \boldsymbol{a'}_1} = \frac{\boldsymbol{a}_1^t C \boldsymbol{a}_2}{\boldsymbol{a'}_1^t C \boldsymbol{a'}_2} = \frac{\boldsymbol{a}_2^t C \boldsymbol{a}_2}{\boldsymbol{a'}_2^t C \boldsymbol{a'}_2} \tag{5.6}$$

We can write these equations directly in terms of the SVD of the fundamental matrix.

$$A' = \begin{pmatrix} \boldsymbol{a'}_1^t \\ \boldsymbol{a'}_2^t \\ \boldsymbol{a'}_3^t \end{pmatrix} = \begin{pmatrix} r & & \\ & s & \\ & & 1 \end{pmatrix} U^t = \begin{pmatrix} r\mathrm{u}_1^t \\ s\mathrm{u}_2^t \\ \mathrm{u}_3^t \end{pmatrix}$$

where u_i^t is the i-th column of U. And

$$A = \begin{pmatrix} \boldsymbol{a}_1^t \\ \boldsymbol{a}_2^t \\ \boldsymbol{a}_3^t \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} V^t = \begin{pmatrix} \mathrm{v}_2^t \\ -\mathrm{v}_1^t \\ \mathrm{v}_3^t \end{pmatrix}$$

where v_i^t is the i-th column of V. From (5.6) we obtain

$$\frac{\mathrm{v}_2^t C \mathrm{v}_2}{r^2 \cdot \mathrm{u}_2^t C \mathrm{u}_1} = \frac{-\mathrm{v}_2^t C \mathrm{v}_1}{rs \cdot \mathrm{u}_1^t C \mathrm{u}_2} = \frac{-\mathrm{v}_1^t C \mathrm{v}_1}{s^2 \cdot \mathrm{u}_2^t C \mathrm{u}_2}$$

Our problem has five degrees of freedom. Each pair of images provides two independent constraints. From three images we can form three pairs which provide three pairs of constraints. This is enough to solve for all the variables in C. However, note that all of the equations are multivariable quadratics in the coefficients of C. This makes the problem quite painful to solve in practise.

The difficulty of such purely algebraic approaches explains why alternative approaches have been explored for self calibration. [10] provides one such alternative. In any case, an algebraic solution can only ever provide the essential first step for a more refined bundle adjustment (error minimization) process.

Bibliography

[1] H.A. Beyer. *Geometric and Radiometric Analysis of a CCD-Camera Based Photogrammetric Close-Range System*. PhD thesis, ETH-Zurich, 1992.

[2] P. Brand, R. Mohr, and Ph. Bobet. Distorsion optique : correction dans un modèle projectif. In *Actes du 9ème Congrès AFCET de Reconnaissance des Formes et Intelligence Artificielle, Paris, France*, pages 87–98, Paris, January 1994.

[3] O. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In G. Sandini, editor, *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italy*, pages 563–578. Springer-Verlag, May 1992.

[4] O. Faugeras. *Three-Dimensional Computer Vision - A Geometric Viewpoint*. Artificial intelligence. M.I.T. Press, Cambridge, MA, 1993.

[5] O. Faugeras. Stratification of three-dimensional vision: Projective, affine and metric representations. *Journal of the Optical Society of America*, 12:465–484, 1995.

[6] J. Forgarty. *Invariant Theory*. Benjamin, New York, USA, 1969.

[7] P. Gros. How to use the cross ratio to compute 3D invariants from two images. In J.L. Mundy, A. Zisserman, and D. Forsyth, editors, *Proceeding of the* DARPA–ESPRIT *workshop on Applications of Invariants in Computer Vision, Azores, Portugal*, Lecture Notes in Computer Science, pages 107–126. Springer-Verlag, 1993.

[8] R. Hartley. In defence of the 8-point algorithm. In *Proceedings of the 5th International Conference on Computer Vision, Cambridge, Massachusetts, USA*, pages 1064–1070, June 1995.

[9] R. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Urbana-Champaign, Illinois, USA*, pages 761–764, 1992.

[10] R.I. Hartley. Euclidean reconstruction from uncalibrated views. In *Proceeding of the* DARPA–ESPRIT *workshop on Applications of Invariants in Computer Vision, Azores, Portugal*, pages 187–202, October 1993.

[11] K. Kanatani. *Geometric Computation for Machine Vision*. Oxford Science Publications, Oxford, 1993.

[12] Q.T. Luong and O. Faugeras. The fundamental matrix: Theory, algorithms and stability analysis. *International Journal of Computer Vision*, 17(1):43–76, 1996.

[13] S.J. Maybank and O.D. Faugeras. A theory of self calibration of a moving camera. *International Journal of Computer Vision*, 8(2):123–151, 1992.

[14] P. Meer, S. Ramakrishna, and R. Lenz. Correspondence of coplanar features through p^2-invariant representations. In *Proceedings of the 12th International Conference on Pattern Recognition, Jerusalem, Israel*, pages A–196–202, 1994.

[15] R. Mohr, B. Boufama, and P. Brand. Understanding positioning from multiple images. *Artificial Intelligence*, (78):213–238, 1995.

[16] T. Moons, L. Van Gool, M. Van Diest, and E. Pauwels. Affine reconstruction from perspective image pairs. In *Proceeding of the* DARPA–ESPRIT *workshop on Applications of Invariants in Computer Vision, Azores, Portugal*, pages 249–266, October 1993.

[17] L. Morin. *Quelques contributions des invariants projectifs à la vision par ordinateur*. PhD thesis, Institut National Polytechnique de Grenoble, January 1993.

[18] J.L. Mundy and A. Zisserman, editors. *Geometric Invariance in Computer Vision*. MIT Press, Cambridge, Massachusetts, USA, 1992.

[19] J.L. Mundy and A. Zisserman, editors. *Proceedings of the Second* ESPRIT - ARPA *Workshop on Applications of Invariance on Computer Vision, Ponta Delagada, Azores, Portugal*, October 1993.

[20] W.H. Press and B.P. Flannery and S.A. Teukolsky and W.T. Vetterling. *Numerical Recipes in C*. Cambridge University Press, 1988.

[21] L. Robert and O. Faugeras. Relative 3d positionning and 3D convex hull computation from a weakly calibrated stereo pair. In *Proceedings of the 4th International Conference on Computer Vision, Berlin, Germany*, pages 540–544, May 1993.

[22] C.A. Rothwell. *Object Recognition Through Invariant Indexing*. Oxford University Press, 1995.

[23] J.G. Semple and G.T. Kneebone. *Algebraic Projective Geometry*. Oxford Science Publication, 1952.

[24] C.C. Slama, editor. *Manual of Photogrammetry, Fourth Edition*. American Society of Photogrammetry and Remote Sensing, Falls Church, Virginia, USA, 1980.

[25] E. Thompson. The projective theory of relative orientation. *Photogrammetria*, 23(1):67–75, 1968.

[26] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization method. *International Journal of Computer Vision*, 9(2):137–154, 1992.

[27] Z. Zhang, R. Deriche, O. Faugeras, and Q.T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. Rapport de recherche 2273, INRIA, May 1994.