Tutorial I: The Algebraic Approach to Invariance

Richard Hartley and Andrew Zisserman

ESPRIT Workshop May 1994

Contents

1 Introduction

An invariant of a geometric configuration is a function of the configuration whose value is unchanged by a particular transformation. For example, the separation of two points is unchanged by a Euclidean transformation (translation and rotation), but not by similarity (translation, rotation and isotropic scaling). Distance is an Euclidean invariant, angle a similarity invariant.

Within the context of vision we are interested in determining invariants under perspective projection of an object into an image. For example, for a planar object the perspective projection between object and image planes is a *projective* transformation. Properties such as intersection, collinearity and tangency are unaffected by a projective transformation, however, invariant *values* can also be computed.

More formally, under a linear transformation of coordinates, $\mathbf{X}' = \mathtt{T}\mathbf{X}$, the invariant $I(\mathbf{X})$ transforms as

$$I(\mathbf{X}') = |\mathtt{T}|^w I(\mathbf{X})$$

and is called a *relative* invariant of weight w. If $w = 0$, the invariant is unchanged under transformations and is called a *scalar* invariant. We will only be interested in scalar invariants in the following.

In general we seek invariance to *projective* transformations, so $\mathtt{T}$ is a general non-singular square matrix acting on homogeneous coordinates. For planar configurations it is 3×3, and for 3D configurations 4×4. The goal is to measure the invariants from a perspective image of the configuration. We write $\mathtt{P}$ for the projection matrix that covers a 3D Euclidean transformation of the object followed by perspective projection onto the image. Affine and similarity invariants (scaled Euclidean) are also important in vision applications, but are not covered here.

1.1 Overview

In order to develop invariants systematically, it is important first to *model* the transformations and projections that occur in imaging. Having established these transformations, and the circumstances under which they apply, we are in a position to derive their invariants.

Part I Describes camera models for projecting from 3D to 2D images, with special cases and decompositions. For *planar* objects the original and image spaces are the same dimension and $\mathtt{P}$ is simply a projective transformation represented by a 3×3 matrix. A description is given of projective transformations and special cases, including Euclidean, similarity, and affine.

For two views of the scene, the fundamental matrix encapsulates the projective properties of the camera pair. The relation between projective cameras and the fundamental matrix is derived. etc.

Part II Examples are given of plane projective invariants, and an an application of these to model based recognition. For 3D objects, the original and image spaces are no longer of the same dimension and $\mathtt{P}$ is a 3×4 matrix mapping 3D homogeneous coordinates onto the image plane. Projective invariants of special 3D structures can be obtained from a single perspective image. A number of examples are given.

Part II

These notes are an introduction only. They are not comprehensive or conventionally referenced. Further reading is included at the end of each section.

Acknowledgements

These notes are drawn from many sources, in particular from the introduction and appendix in [5], written with Joe Mundy. As ever we are very grateful for discussions with Joe Mundy, David Forsyth, Charlie Rothwell, Paul Beardsley and ???. Charlie Rothwell provided most of the images and implementation. Financial support was provided by Esprit Project 6448 'VIVA' and ...

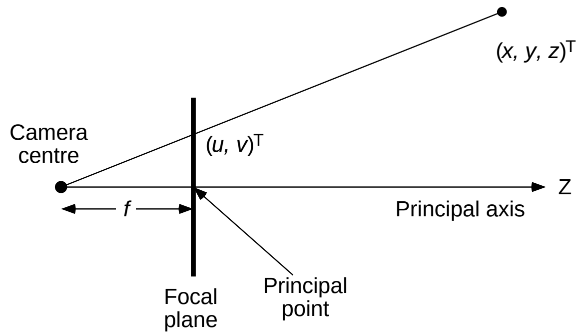


Figure 1: *Pinhole camera geometry*

Part I
Transformations and Projections

2 Camera Models

The basic pinhole model We discuss here the pinhole or central projection camera model. We consider the central projection of points in space onto a plane. Let the centre of projection be the origin of a Euclidean coordinate system, and consider the plane $z = f$, henceforth called the focal plane. Under the pinhole camera model, a point in space with coordinates $\mathbf{x} = [x, y, z]^\top$ mapped to the point in the focal plane where a line joining the point $\mathbf{x}$ to the centre of projection meets the focal plane. This is shown in Fig 2. By similar triangles, one quickly computes that the point $[x, y, z]^\top$ is mapped to the point $[fx/z, fy/z, f]^\top$ on the focal plane. Ignoring the final coordinate, we see that

$$[x, y, z]^\top \mapsto [fx/z, fy/z]^\top \tag{1}$$

describes the central projection mapping. This is a mapping from Euclidean space R^3 to R^2.

The centre of projection is often called the *camera centre*. The line from the camera centre perpendicular to the focal plane is called the *principal axis* of the camera, and the point where the principal axis meets the focal plane is called the *principal point*.

Homogeneous Coordinates It is convenient to express points in space and points in the image plane in homogeneous coordinates. A point in 3-space R^3 is expressed in homogeneous coordinates by a 4-vector. Specifically, the homogeneous vector $[x, y, z, t]^\top$ with $t \neq 0$ represents the point $[x/t, y/t, z/t]^\top$ of R^3 in non-homogeneous coordinates. Similarly, a homogeneous vector $[u, v, w]^\top$ represents the point $[u/w, v/w]^\top$ in R^2. One sees immediately that two homogeneous vectors that differ by a constant non-zero factor represent the same point. Consequently, two homogeneous vectors differing by a non-zero constant factor are considered to be equivalent. We may write $\mathbf{x} \approx \mathbf{x}'$ to express this equivalence of homogeneous vectors. However, this notation quickly becomes tedious, and so usually the equivalence of two homogeneous vectors is expressed using an equality sign. Thus we write $\mathbf{x} = \mathbf{x}'$ to mean that the two vectors are equal up to a multiplicative factor.

The plane at infinity It was seen in the previous paragraph that a homogeneous vector $\mathbf{x} = [x, y, z, t]^\top$ represents a point in R^3 if and only if $t \neq 0$. The set of all non-zero homogeneous 4-vectors form the projective 3-space P^3. The set of points $\mathbf{x} = [x, y, z, 0]^\top$ form a plane consisting of points not in R^3. This is referred to as the plane at infinity. Thus, projective 3-space P^3 is made up of R^3 plus the plane at infinity. A point in R^3 may be conveniently expressed in homogeneous coordinates as $[x, y, z, 1]^\top$.

In the same way, projective 2-space P^2 is made up of R^2 plus a line at infinity, consisting of points $[u, v, w]^\top$ in homogeneous coordinates with $w = 0$.

Projective geometry is the study of the projective space P^n. In projective geometry it is usual not to distinguish the plane (or line) at infinity from any other plane. Thus, all points, whether infinite of finite are created equal. In the area of computer vision dealing with points in space and projections of these points by pinhole cameras, we also deal with homogeneous vectors. However, in computer vision it is often appropriate to distinguish the plane at infinity and treat it differently from other planes. For instance, no person ever managed to photograph a point at infinity, nor did anyone ever manage to place a camera on the plane at infinity. Certain concepts such as front and back of the camera and affine and Euclidean reconstruction and invariants make no sense without considering the plane at infinity to be distinguished.

Central projection using homogeneous coordinates The purpose of expressing points in homogeneous coordinates is that central projection is very simply expressed as a linear mapping in homogeneous coordinates. In particular, the expression (1) may be written in terms of matrix multiplication as

$$[x, y, z, 1]^\top \mapsto [fx, fy, z]^\top = \begin{bmatrix} f & & & 0 \\ & f & & 0 \\ & & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{2}$$

The matrix in this expression may be written as $\mathrm{diag}(f, f, 1)[I \mid \mathbf{0}]$ where $\mathrm{diag}(f, f, 1)$ is a diagonal matrix and $[I \mid \mathbf{0}]^\top$ represents a matrix divided up into a 3×3 block (the identity matrix) plus a column vector, here the zero vector. If the image of a point $\mathbf{x}$ under central projection is $\mathbf{u}$ then we see that

$$\mathbf{u} = \mathrm{diag}(f, f, 1)[I \mid \mathbf{0}] \ .$$

Principal point offset The expression (1) assumed that the origin of coordinates in the image plane is at the principal point. In practice, the principal point may not be accurately known. In general, therefore, we will have a mapping

$$[x, y, z]^\top \mapsto [fx/z + p_u, fy/z + p_v]^\top$$

where $(p_u, p_v)^\top$ are the coordinates of the principal point, otherwise known as the *principal point offset*. This equation may be expressed conveniently in homogeneous coordinates as

$$\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \mapsto \begin{bmatrix} fx + p_u \\ fy + p_v \\ z \end{bmatrix} = \begin{bmatrix} f & & p_u & 0 \\ & f & p_v & 0 \\ & & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{3}$$

Camera Rotation and Translation So far, it has been assumed that the camera is located at the origin of a Euclidean coordinate system with the principal axis of the camera pointing straight down the z-axis. Such a coordinate system may be called the *camera coordinate frame.* In general, however, points in space will be expressed in terms of a different Euclidean coordinate frame, known as the *world* coordinate frame. The two coordinate frames are related via a rotation and a translation. If $\mathbf{x}$ is a non-homogeneous vector representing the coordinates of a point in the world coordinate frame, and $\mathbf{x}'$ represents the same point in the camera coordinate frame, then we may write $\mathbf{x}' = R(\mathbf{x} - \mathbf{c})$ where $\mathbf{c}$ represents the coordinates of the camera centre in the world coordinate frame, and R is a 3×3 rotation matrix representing the orientation of the camera coordinate frame. This equation may be written in homogeneous coordinates as

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} R & -R\mathbf{c} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

Putting this together with the equation 2 leads to the formula

$$\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \mapsto \begin{bmatrix} f & & p_u \\ & f & p_v \\ & & 1 \end{bmatrix} R[I \mid -\mathbf{c}] \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{4}$$

where

- f is the focal length of the camera.
- $(p_u, p_v)^\top$ are the image coordinates of the principal point.
- R is the rotation of the camera.
- $\mathbf{c}$ is the location of the camera centre.

Writing

$$K = \begin{bmatrix} f & & p_u \\ & f & p_v \\ & & 1 \end{bmatrix} \tag{5}$$

we see that image of a point $\mathbf{x}$ under a pinhole camera mapping is

$$\mathbf{u} = KR[I \mid -\mathbf{c}]\mathbf{x} \tag{6}$$

This is the general mapping given by a pinhole camera. One sees that a general pinhole camera has 9 degrees of freedom.

CCD cameras The pinhole camera model just derived assumes that both object coordinates (that is the 3D world coordinates) and image coordinates are Euclidean coordinates having equal scales in all axial directions. In the case of CCD cameras, there is the additional possibility of having unsquare pixels. If image coordinates are measured in pixels, then this has the extra effect of introducing unequal scale factors in each direction. In particular if the number of pixels per unit distance in image coordinates are m_u and m_v in the u and v directions, then the cameras transformation from world coordinates to pixel coordinates is obtained by multiplying (5) on the left by an extra factor $\mathrm{diag}(m_1, m_2, 1)$. Thus, the general form of the calibration matrix of a CCD camera is

$$K = \begin{bmatrix} k_u & & p_u \\ & k_v & p_v \\ & & 1 \end{bmatrix} \tag{7}$$

where $k_u = fm_u$ and $k_v = fm_v$ represent the focal length of the camera in terms of pixel dimensions in the u and v direction respectively. Similarly, (p_u, p_v) are the pixel coordinates of the principal point. A CCD camera thus has 10 degrees of freedom.

General Projective Camera For simplicity and added generality, we can consider a calibration matrix of the form

$$K = \begin{bmatrix} k_u & s & p_u \\ & k_v & p_v \\ & & 1 \end{bmatrix} \ . \tag{8}$$

The added parameter s is referred to as the *skew* parameter. The skew parameter will be zero for most normal cameras. However, in certain unusual instances it can take non-zero values. In a CCD camera if the pixel elements in the CCD array are skewed so that the u and v axes are not perpendicular, then skewing of the image can result. This is admittedly very unlikely to happen. The CCD camera model assumes that the image has been stretched by different amounts in the two axial directions. If on the other hand the image is stretched in a non-axial direction, then skewing results. To see this, consider what happens to a pair of axes if the image is stretched in a diagonal direction : the axes do not remain perpendicular. Skewing may occur if images taken by a pinhole camera (such as an ordinary film camera) is subsequently magnified. If the axis of the magnifying lens is not perpendicular to the film plane or the new image plane, then the image will be skewed. In all of these cases, the effect of skew will be small, so generally the parameter s will be very small compared with k_u.

A camera with camera matrix $P = KR[I \mid -\mathbf{c}]$ for which the calibration matrix K is of the form (8) will be called a *projective* camera. A projective camera has 11 degrees of freedom. This is the same number of degrees of freedom as a 3×4 matrix, defined up to an arbitrary scale. Any general 3×3 matrix M may be decomposed as a product $M = KR$ where K is upper triangular and R is a rotation matrix. Thus, the class of projective cameras with camera centre at a finite point corresponds to the class of matrices of the form $P = [M \mid -M\mathbf{c}]$. This is a general 3×4 matrix with the sole restriction that the left-hand 3×3 block is non-singular. We may further extend the class of camera matrices to include cameras at infinity. A general projective camera is one which is represented by an arbitrary 3×4 matrix of rank 3.

In summary, we may distinguish the following hierarchy of camera models.

General Projective Camera. The general projective camera is one for which the object-space to image-space transformation is represented by the mapping $\mathbf{u} = P\mathbf{x}$ in homogeneous coordinates, where P is an arbitrary 3×4 matrix of rank 3. Matrix P is defined only up to a non-zero multiplicative factor:

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{9}$$

Finite Projective Camera. This terminology may be used to describe a projective camera for which the camera centre is not at infinity. The camera matrix P may be written in the form $P = [M \mid -M\mathbf{c}]$ where $\mathbf{c}$ is the camera centre and M is non-singular. Using the QR decomposition M can be decomposed as $M = KR$, where R is a rotation and K is an upper triangular matrix with positive diagonal entries, called the *calibration matrix*.

Zero-skew camera This is a finite projective camera with zero skew, such as a CCD camera. Hence the calibration matrix has the form (7).

Pinhole camera A finite projective camera with calibration matrix of the form (4) having zero skew and equal magnification in each direction.

Calibrated Camera. A camera with matrix of the form $P = [R \mid -R\mathbf{c}]$. In other words, the calibration matrix in assumed to be the identity. It is not implied, of course that the calibration matrix really is the identity matrix, but rather that the effect of the calibration matrix K, once it is known, may be removed, and that one may without loss of generality assume that $K = I$. Accordingly, the term *calibrated camera* will be used in these notes to mean that $K = I$.

All these different camera models have been considered in the literature. The terminology proposed here is certainly not standardized. For instance, the term pinhole camera has been used to denote more general camera models, even the general projective camera model. For the most part, in these notes we will consider either general or finite projective cameras. When we talk of camera calibration, or cameras with specific calibration, however, it is implicit that we are talking of finite projective cameras, since there is no natural way to define the calibration matrix of a camera at infinity.

Cameras with centres at infinity form a different hierarchy, which shall not be discussed further in these notes.

3 Plane to plane transformations

In the case that the 3D points $\mathbf{x}$ lie on a plane, then without loss of generality, the z coordinate of $\mathbf{x}$ can be chosen as zero. The general projection between world and image, camera (9), reduces to a 3×3 matrix. This is a *plane projective transformation* (projectivity or collineation) and covers the composed effects of a Euclidean transformation and perspective projection. See figure 2.

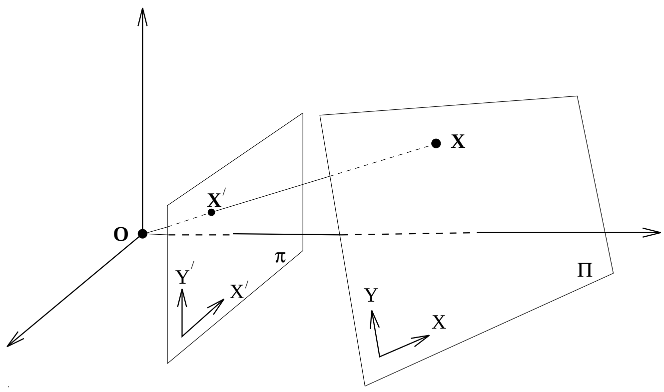


Figure 2: A perspective transformation between two planes: a world plane, Π, and an image plane, π, with optical center $\mathbf{O}$. The projection is modelled as a linear transformation, $\mathbf{X}' = \mathtt{T}\mathbf{X}$, on homogeneous point coordinates.

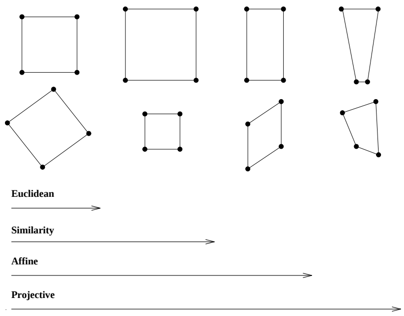


Figure 3: THIS FIGURE WILL BE INCLUDED IN THE TABLE. The hierarchy of plane to plane transformations. These range from Euclidean, where only translations and rotations occur, to projective where the square can be transformed to any arbitrary quadrilateral (provided no three points are collinear). Typical invariants to these transformations are given in table 1.

3.1 Plane projectivities

A plane projectivity is a linear transformation of homogeneous 3-vectors represented by a 3×3 matrix, $\mathbf{X}' = \mathtt{T}\mathbf{X}$,

$$\begin{bmatrix} X_1' \\ X_2' \\ X_3' \end{bmatrix} = \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}$$

where

$$X' = X_1'/X_3' \qquad Y' = X_2'/X_3'$$

Eight parameters are required to define the 2D projective transformation matrix, $\mathtt{T}$, up to an arbitrary scale factor. Special cases of projective transformations are illustrated in figure 3 and their form given in table 1.

Group	DOF	Matrix	Invariant properties
projective	8	$\begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{bmatrix}$	concurrency and collinearity, **order of contact**: intersection (1 pt contact); tangency (2 pt contact); inflections (3 pt contact with line); tangent discontinuities and cusps. cross-ratio (ratio of ratio of lengths).
affine	6	$\begin{bmatrix} a_{11} & a_{12} & t_{13} \\ a_{21} & a_{22} & t_{23} \\ 0 & 0 & 1 \end{bmatrix}$	parallelism, ratio of areas, ratio of lengths on collinear or parallel lines (e.g. midpoints) linear combinations of vectors
similarity	4	$s \begin{bmatrix} r_{11} & r_{12} & t_x \\ r_{21} & r_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix}$	ratio of lengths, angle
Euclidean	3	$\begin{bmatrix} r_{11} & r_{12} & t_x \\ r_{21} & r_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix}$	length, area

Table 1: Geometric properties invariant to commonly occurring plane transformations. Transformations lower in the table inherit the invariants of those above, but the converse is not true. The matrix $\mathtt{A} = [a_{ij}]$ is an invertible 2×2 matrix, $\mathtt{R} = [r_{ij}]$ is a 2D rotation matrix, and (t_x, t_y) a 2D translation.

3.2 Homogeneous line representation

The equation of a line in homogeneous coordinates is given by

$$L_1 X_1 + L_2 X_2 + L_3 X_3 = 0 \tag{10}$$

or $\mathbf{L}.\mathbf{X} = 0$ where $\mathbf{L} = (L_1, L_2, L_3)^\top$. Thus a line is represented in homogeneous form by three numbers (L_1, L_2, L_3). Note that, as in the representation of a point, only the ratio of these numbers is relevant, since multiplying equation (10) by a scalar λ has no effect on the line.

Under the point transformation $\mathbf{X}' = \mathtt{T}\mathbf{X}$ a line transforms as $\mathbf{L}' = \mathtt{T}^{-t}\mathbf{L}$. Thus, lines in the projective plane transform linearly, just as points, but the corresponding transformation matrix is the transpose of the inverse of the matrix defining the point transformation.

3.3 Duality

Points and lines have a symmetric role in the projective geometry of the plane. For example, it can be shown that the line, $\mathbf{l}$, through two points $\mathbf{p}, \mathbf{q}$ is given by the vector product of the points

$\mathbf{l} = \mathbf{p} \times \mathbf{q}$, and the point, $\mathbf{p}$, at which two lines $\mathbf{l}, \mathbf{m}$ intersect is given by the vector product of the lines $\mathbf{p} = \mathbf{l} \times \mathbf{m}$. This is an example of the "Principle of Duality", which says that any statement or theorem involving points or lines still holds with the word point replaced by line and vice-versa.

3.4 Projective transformations of conics

The equation of a conic in non-homogeneous coordinates is

$$AX^2 + BXY + CY^2 + DX + EY + F = 0$$

"Homogenising" this gives:

$$A{X_1}^2 + BX_1X_2 + C{X_2}^2 + DX_1X_3 + EX_2X_3 + F{X_3}^2 = 0$$

or in matrix form $\mathbf{X}^\top \mathtt{C}\mathbf{X} = 0$ The conic coefficient matrix $\mathtt{C}$ is given by

$$\mathtt{C} = \begin{bmatrix} A & B/2 & D/2 \\ B/2 & C & E/2 \\ D/2 & E/2 & F \end{bmatrix} .$$

As in the case of points and lines only the ratio of the matrix elements are important, since multiplying $\mathtt{C}$ by a non-zero scalar does not affect the above equations. Under the point transformation $\mathbf{X}' = \mathtt{T}\mathbf{X}$, a conic transforms as $\mathtt{C}' = \mathtt{T}^{-\top}\mathtt{C}\mathtt{T}^{-1}$ where $\mathbf{X}^\top\mathtt{C}\mathbf{X} = 0$ is transformed to $\mathbf{X}'^\top\mathtt{C}'\mathbf{X}' = 0$.

4 The Fundamental Matrix

In 1982, Longuet-Higgins ([19]) gave a solution to the relative placement problem (computation of relative camera placement from image correspondences) by introducing what became known as the *essential matrix*, denoted Q. He showed that for a pair of calibrated cameras with matrices $P = [I \mid \mathbf{0}]$ and $P' = [R \mid -R\mathbf{c}]$ there exists a matrix Q with the following property. If $\mathbf{u} \leftrightarrow \mathbf{u}'$ are a pair of corresponding points in the two images, then $\mathbf{u}'^\top Q\mathbf{u} = 0$.

To consider this further, we need a new terminology. Let $\mathbf{t}$ be a 3-vector and M be a 3×3 matrix. We denote by $\mathbf{t} \times M$ the matrix formed by taking the cross-product of $\mathbf{t}$ with the columns of M separately. Similarly, $M \times \mathbf{t}$ is the matrix formed by taking the cross product of each rows of M with $\mathbf{t}$ separately. If $\mathbf{t} = [t_x, t_y, t_z]$, then we define a matrix

$$[\mathbf{t}]_\times = \begin{bmatrix} 0 & -t_x & t_y \\ t_x & 0 & -t_z \\ -t_y & t_z & 0 \end{bmatrix} . \qquad (11)$$

One quickly verifies that

$$\mathbf{t} \times M = [\mathbf{t}]_\times M$$

and

$$M \times \mathbf{t} = M[\mathbf{t}]_\times .$$

Longuet-Higgins showed (effectively) that the essential matrix Q corresponding to a pair of camera matrices $[I \mid \mathbf{0}]$ and $[R \mid -R\mathbf{c}] = [R \mid \mathbf{t}]$ is the matrix

$$Q = \mathbf{t} \times R .$$

Longuet-Higgins idea is to use sufficiently many point matches $\mathbf{u}_i \leftrightarrow \mathbf{u}'_i$ to determine Q using the relation $\mathbf{u}'^{\top} Q\mathbf{u} = 0$. With at least 8 point matches, Q may be determined using linear techniques. Subsequently, the matrix Q may be factored as $Q = \mathbf{t} \times R$ to find the two camera matrices.

4.1 Generalization to Projective Cameras

Much of the work of Longuet-Higgins may be generalized to projective cameras. To conform with current terminology, we will define a matrix, called the *fundamental matrix*, denoted F, which is associated with a pair of camera matrices. In the case where the camera matrices correspond to calibrated cameras, F will be identical with the essential matrix Q of Longuet-Higgins.

Summary of Properties of the Fundamental Matrix. We list some of the properties of the Fundamental matrix.

Proposition 4.1.

1. *F is a 3×3 matrix of rank 2.*
2. *If $\mathbf{u} \leftrightarrow \mathbf{u}'$ are a pair of matching points, then $\mathbf{u}'^{\top} F\mathbf{u} = 0$*
3. *If F is the fundamental matrix for a pair of cameras (J, J') then $F^{\top}$ is the fundamental matrix for the pair (J', J).*
4. *If $\mathbf{p}$ and $\mathbf{p}'$ are the two epipoles, then $\mathbf{p}'^{\top} F = F\mathbf{p} = 0$.*
5. *$F\mathbf{u}$ is the epipolar line in the second image corresponding to point $\mathbf{u}$ in the first image.*
6. *F factors as a product $F = \mathbf{p}'^{\top} \times M$, where M is non-singular.*

Formulas for the Fundamental Matrix The Fundamental matrix is uniquely determined by a pair of camera matrices. Thus, suppose that P and P' are a pair of camera matrices with the sole restriction that they do not have the same centre.

Proposition 4.2.

1. *F is the unique non-zero matrix such that $P'^{\top} F P$ is skew-symmetric.*
2. *If $P = [M \mid -M\mathbf{c}]$ and $P' = [M' \mid -M'\mathbf{c}']$ then*

$$\begin{aligned} F &= [M'(\mathbf{c}' - \mathbf{c})]_{\times}(M'M^{-1}) \\ &= M'^{-\top}[\mathbf{c}' - \mathbf{c}]_{\times} M^{-1} \\ &= (M'M^{-1})^{-\top}[M(\mathbf{c}' - \mathbf{c})]_{\times} \end{aligned}$$

3. *If $P = [I \mid 0]$ and $P' = [M' \mid \mathbf{t}]$ then $F = \mathbf{t} \times M'$*

Determination of the Camera Matrices from the Fundamental Matrix. In the last paragraph, it was stated that the fundamental matrix is completely determined by a pair of camera matrices. The converse is not true, as is summarized in the following set of results.

Proposition 4.3. *Let H represent a 3D projective transform.*

1. *Camera matrix pairs (P, P') and $(PH^{-1}, P'H^{-1})$ have the same fundamental matrix.*
2. *F determines the pair of camera matrices up to a projective transform, H.*
3. *Given F, we can always find a pair of camera matrices $P = [I \mid \mathbf{0}]$ and $P' = [M \mid \mathbf{t}]$ that realize F.*

4.2 Proofs of properties of the Fundamental Matrix

The following section will be devoted to proving some of the properties of the fundamental matrix stated in the previous section.

Let P and P' be a pair of camera matrices and let $\mathbf{u}' \leftrightarrow \mathbf{u}$ be a pair of matched points. This means that there is a point $\mathbf{x}$ in space such that $\mathbf{u} = P\mathbf{x}$ and $\mathbf{u}' = P'\mathbf{x}$. We seek a matrix F (the fundamental matrix) such that $\mathbf{u}'^\top F\mathbf{u} = 0$. Expressing this in terms of F leads to the equation

$$\mathbf{x}^\top P'^\top F P\mathbf{x} = 0$$

This equation must hold for any point $\mathbf{x}$ in space, since any such point gives rise to a pair of matched points $\mathbf{u} = P\mathbf{x}$ and $\mathbf{u}' = P'\mathbf{x}$. However, a matrix A satisfies the equation $\mathbf{x}^\top A\mathbf{x} = 0$ for all $\mathbf{x}$ if and only if A is skew-symmetric. Consequently, we have proven the following result

Proposition 4.4. *Given a pair of camera matrices P and P', then a matrix F satisfies $\mathbf{u}'^\top F\mathbf{u} = 0$ for all possible matched points $\mathbf{u} \leftrightarrow \mathbf{u}'$ in the images taken by the two cameras, if and only if $P'^\top FP$ is skew-symmetric.*

Under this condition we say that the pair (P, P') is a *realisation* of the fundamental matrix F.

Two questions arise from this proposition. To what extent is the matrix F determined by the two matrices P and P', and secondly, to what extent are P and P' determined by the matrix F. It was stated in Proposition 4.2 that this condition uniquely determines F. This will be shown later.

We first show that given matrix F, the matrices P and P' are not uniquely determined. Specifically, if H is a non-singular 4×4 matrix, and $P'^\top FP$ is skew-symmetric, then so is $H^\top P'^\top FPH$. This shows that (P, P') and $(P'H, PH)$ are both realizations of the matrix F. It will be shown later that this is the only ambiguity in the realization of a fundamental matrix.

Projective transform of a reconstruction. This may be seen in another way as follows. If $\mathbf{u} = P\mathbf{x}$ and H is a non-singular 4×4 matrix representing any projective transformation then replacing the camera P by PH^{-1} and the point $\mathbf{x}$ by $H\mathbf{x}$, we see that $\mathbf{u} = (PH^{-1})H\mathbf{x}$. Thus, the image point $\mathbf{u}$ is unchanged by this transformation. This shows that the camera matrices can not be determined unambiguously by a set of image correspondences, since an arbitrary projective transformation H applied to the scene and the cameras in this way does not result in any change in the images. Thus, neither the scene, nor the camera placement may be determined mor accurately than up to an unknown projective transformation. We speak of a projective transformation H being applied to a camera matrix P to mean that P is replaced by PH^{-1}.

Normal form. Given a matrix F, we have just seen that there exist a large family of camera matrix pairs, (P, P') that make a realization of F. Next, we will be interested in certain normal form realizations. The first *normal form* will be one in which one of the cameras has matrix $[I \mid \mathbf{0}]$ and the other camera has camera matrix $[M \mid -M\mathbf{c}]$, meaning that the camera centre is not at infinity. In fact, this can be done with any number of cameras, as the following proposition shows.

Proposition 4.5. *Given camera matrices P_i for $i = 0, \ldots, N$, there exists a 4×4 matrix H such that $P_0 H^{-1} = [I \mid \mathbf{0}]$ and for all $i = 1, \ldots, N$ the camera centre for the matrix $P_i H^{-1}$ does not lie at infinity.*

Proof. Since P_0 has rank 3, it may be supplemented by one extra row to form a non-singular matrix, which we will denote by H. We see immediately that $P_0 H^{-1} = [I \mid \mathbf{0}]$. Multiplying each of the other matrices by H^{-1} gives matrices $P_i H^{-1} = P'_i$. If the centre of each camera P'_i lies at a finite point, then we are done. Otherwise, we must apply a further transformation. Let $\mathbf{c}'_i$ be the centre of the camera P'_i. Thus, $P'_i \mathbf{c}'_i = 0$. Now, we select a plane that contains none of the points $\mathbf{c}'_i$, and which furthermore does not pass through the point $[0, 0, 0, 1]^\top$. Such a plane may be represented as $[\mathbf{v}^\top 1]^\top$, where $\mathbf{v}$ is a 3-vector. The condition that the plane does not pass through any of the points $\mathbf{c}'_i$ means that $[\mathbf{v}^\top 1]\mathbf{c}'_i \neq 0$ for any i. Letting $H' = \begin{bmatrix} I & 0 \\ \mathbf{v}^\top & 1 \end{bmatrix}$ this implies that $H'\mathbf{c}'_i$ is not a point at infinity. However, $H'\mathbf{c}'_i$ is the centre of the camera $P'_i H'^{-1}$. One verifies further that $[I \mid \mathbf{0}]H'^{-1} = [I \mid \mathbf{0}]$. Thus, transforming each P'_i by H'^{-1} gives the required set of camera matrices. □

This is a particularly convenient normal form for many applications. It may be easily verified that a matrix of the type

$$H = \begin{bmatrix} I & 0 \\ \mathbf{v}^\top & 1 \end{bmatrix}$$

is the general form for a 4×4 matrix satisfying the condition $[I \mid \mathbf{0}]H^{-1} = [I \mid \mathbf{0}]$. The choice of different vectors $\mathbf{v}$ correspond to different choices of the plane at infinity, since a point is mapped to infinity by the transformation H if and only if it lies on the plane represented by $[\mathbf{v}^\top 1]^\top$.

Now, we may show that the two matrices P and P' uniquely determine F as long as the camera centres for P and P' are not the same. Thus, suppose that $P'^\top F P$ is skew-symmetric. There is a matrix H such that $PH^{-1} = [I \mid \mathbf{0}]$ and $P'H^{-1} = [M \mid -M\mathbf{c}]$, where $\mathbf{c} \neq \mathbf{0}$ and M is non-singular. Now applying $H^{-\top}$ and H^{-1} to the left and right sides of $P'^\top F P$, we see that

$$\begin{aligned} H^{-\top} P'^\top F P H^{-1} &= \begin{bmatrix} M^\top \\ -\mathbf{c}^\top M^\top \end{bmatrix} F [I \mid \mathbf{0}] \\ &= \begin{bmatrix} M^\top F & 0 \\ -\mathbf{c}^\top M^\top F & 0 \end{bmatrix} \end{aligned}$$

is skew-symmetric. This implies that $M^\top F$ is skew-symmetric and $\mathbf{c}^\top (M^\top F) = 0$. Since a 3×3 skew-symmetric matrix is determined by its kernel (in this case $\mathbf{c}$), it follows that $M^\top F = [\mathbf{c}]_\times$. Finally, therefore, $F = M^{-\top}[\mathbf{c}]_\times = M^{-\top} \times \mathbf{c}$. Thus, the matrix F is uniquely determined by P and P'.

It resulted from the previous proof that the matrix $M^{-\top} \times \mathbf{c}$ is the fundamental matrix corresponding to a pair of camera matrices $[I \mid \mathbf{0}]$ and $[M \mid -M\mathbf{c}]$, where M is non-singular. One may remove the restriction that M is non-singular with the following result.

Proposition 4.6. *The fundamental matrix corresponding to a pair of camera matrices $P = [I \mid \mathbf{0}]$ and $P' = [M \mid \mathbf{t}]$ is $F = \mathbf{t} \times M$.*

Proof. One simply verifies that

$$P'^{\top} F P = \begin{bmatrix} M^{\top} \\ \mathbf{t}^{\top} \end{bmatrix} [\mathbf{t}]_{\times} M [I \mid \mathbf{0}] = \begin{bmatrix} M^{\top}[\mathbf{t}]_{\times} M & 0 \\ 0 & 0 \end{bmatrix}$$

is skew symmetric. □

This result may be used to compute the fundamental matrix for any pair of matrices by transforming the two camera matrices to the form given in the proposition by applying an appropriate projective transformation H.

Factorization of the fundamental matrix : Conversely, given a fundamental matrix, it is possible to determine a pair of camera matrices that give rise to that fundamental matrix by applying Proposition 4.6.

Suppose that the singular value decomposition ([1]) of F is given by $F = UDV^{\top}$, where D is the diagonal matrix $D = \mathrm{diag}(r, s, 0)$. The following factorization of F may now be verified by inspection.

$$F = SM \ ; \quad S = UZU^{\top} \ ; \quad M = UE\mathrm{diag}(r, s, \alpha)V^{\top}$$

where

$$E = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad ; \quad Z = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

and α is an arbitrary number. The matrix S is skew-symmetric, $S = [\mathbf{t}]_{\times}$.

By factoring F in this way, as $F = \mathbf{t} \times M$, we can then apply Proposition 4.6 to get a pair of cameras matrices $P = [I \mid \mathbf{0}]$ and $P' = [M \mid \mathbf{t}]$ that correspond to F.

Uniqueness of Camera Matrix up to Projective Transformation. We are almost ready to prove that the fundamental matrix determines the two camera matrices up to a projective transformation. First, however, we need one more lemma.

Lemma 4.7. *Let the rank 2 matrix F factor in two different ways as $F = \mathbf{t} \times M = \mathbf{t}' \times M'$. Then $\mathbf{t} = \mathbf{t}'$ and $M' = M + \mathbf{t}\mathbf{a}^{\top}$ for some vector $\mathbf{a}$.*

Proof. First, note that $\mathbf{t}F = \mathbf{t}[\mathbf{t}]_{\times} M = 0$, and similarly, $\mathbf{t}'F = 0$. Since F has rank 2, it follows that $\mathbf{t} = \mathbf{t}'$ as required. Next, from $[\mathbf{t}]_{\times} M = [\mathbf{t}]_{\times} M' = F$ it follows that $[\mathbf{t}]_{\times}(M' - M) = 0$, and so $M' - M = \mathbf{t}\mathbf{a}^{\top}$ for some $\mathbf{a}$. Hence, $M' = M + \mathbf{t}\mathbf{a}^{\top}$ as required. □

We now answer the question when two pairs of camera matrices may correspond to the same fundamental matrix.

Theorem 4.8. *Let (P_1, P_1') and (P_2, P_2') be two pairs of camera transforms. Then (P_1, P_1') and (P_2, P_2') correspond to the same fundamental matrix F if and only if there exists a 4×4 non-singular matrix H^{-1} such that $P_1 H^{-1} = P_2$ and $P_1' H^{-1} = P_2'$.*

Proof. The *if* part of this theorem has already been proven, so we turn to the *only if* part. Since each of the matrices P_1 and P_2 has rank 3, we can multiply them (on the right) by suitable matrices H_1 and H_2 to transform them each to the matrix $[I \mid \mathbf{0}]$. If the matrices P'_1 and P'_2 are also multiplied by H_1 and H_2 respectively, then the fundamental matrix corresponding to the camera matrix pairs are unchanged, as seen previously. Thus, we have reduced the problem to the case where $P_1 = P_2 = [I \mid \mathbf{0}]$.

Suppose therefore, that $P_1 = P_2 = [I \mid \mathbf{0}]$ and that $P'_1 = [M'_1 \mid \mathbf{t}'_1]$ and $P'_2 = [M'_2 \mid \mathbf{t}'_2]$. By proposition 4.6 we have $F = [\mathbf{t}'_1]_\times M'_1 = [\mathbf{t}'_2]_\times M'_2$. According to Lemma 4.7 this implies that $\mathbf{t}'_1 = \mathbf{t}'_2 = \mathbf{t}$ and that $M'_2 = M'_1 + \mathbf{t}\mathbf{a}^\top$ for some vector $\mathbf{a}$. Let H^{-1} be the matrix $\begin{bmatrix} I & 0 \\ \mathbf{a}^\top & 1 \end{bmatrix}$. Then one verifies that $[I \mid \mathbf{0}] = [I \mid \mathbf{0}]H^{-1}$, so $P_2 = P_1 H^{-1}$. Furthermore, $P'_1 H^{-1} = [M'_1 \mid \mathbf{t}]H^{-1} = [M'_1 + \mathbf{t}\mathbf{a}^\top \mid \mathbf{t}] = [M'_2 \mid \mathbf{t}] = P'_2$. Thus H^{-1} is the matrix required for the conclusion of theorem 4.8. □

4.3 Point set reconstruction

Given a pair of camera matrices P and P' and a pair of matched points $\mathbf{u} \leftrightarrow \mathbf{u}'$ it is evident that the space point $\mathbf{x}$ that gives rise to the two matching image points is uniquely defined, and may be obtained by intersecting two rays from the camera centres. Here is a simple way of computing the point $\mathbf{x}$.

Suppose that the fundamental matrix factors as as $F = \mathbf{t}' \times M'$, and let $P = [I \mid \mathbf{0}]$ and $P' = [M' \mid \mathbf{t}']$ be a realization of the matrix F. Let $\mathbf{u} \leftrightarrow \mathbf{u}'$ be a pair of matched points in the two images. We wish to find a point $\mathbf{x}$ in space such that $\mathbf{u} = P\mathbf{x}$ and $\mathbf{u}' = P'\mathbf{x}$. From the relation $\mathbf{u}'^\top F\mathbf{u} = \mathbf{u}'^\top [\mathbf{t}']_\times M'\mathbf{u} = \mathbf{u}'^\top(\mathbf{t}' \times M'\mathbf{u}) = 0$, it follows that $\mathbf{u}'$, $M'\mathbf{u}$ and $\mathbf{t}'$ are linearly dependent. If in particular $M'\mathbf{u} = \beta\mathbf{u}' - \alpha\mathbf{t}'$ then we define the corresponding object space point $\mathbf{x}$ to be the point $\begin{bmatrix} \mathbf{u} \\ \alpha \end{bmatrix}$. It is now easily verified that $P\mathbf{x} = [I \mid \mathbf{0}]\mathbf{x} = \mathbf{u}$ and $P'\mathbf{x} = [M' \mid \mathbf{t}']\mathbf{x} = M'\mathbf{u} + \alpha\mathbf{t}' = \mathbf{u}'$. This verifies that the given values of P, P' and $\mathbf{x}_i$ constitute a projective reconstruction of the data.

As shown, $\mathbf{x}$ is determined by the two camera matrices P and P' and the matched points $\mathbf{u} \leftrightarrow \mathbf{u}'$. If we choose a different pair of camera matrices PH and $P'H$ realizing the same fundamental matrix F, then in order to preserve the same pair of matched image points, the point $\mathbf{x}$ must be replaced by $H^{-1}\mathbf{x}$. Thus, changing to a different realization of F results in a projective transformation (namely H^{-1}) of the scene. This proves the following theorem

Theorem 4.9. (Faugeras [5], Hartley et al. [10]) *Given a set of image correspondences $\{\mathbf{u}_i\} \leftrightarrow \{\mathbf{u}'_i\}$ sufficient to determine the fundamental matrix, the corresponding object space coordinates $\{\mathbf{x}_i\}$ may be computed up to a collineation of projective 3-space $\mathcal{P}^3$.*

Further Reading

For methods of computing the Fundamental Matrix, see the work of Faugeras, Luong et al, as well as the original work of Longuet-Higgins [19, 6, 20]. For iterative methods of projective reconstruction, see [5, 22, 13]. A more geometric approach to reconstruction is taken by Ponce et.al ([23]) and Shashua ([30, 31]). Reconstruction from lines instead of points is considered in [15].

5 Self Calibration

So far, we have considered invariants that can be computed from two or more views with arbitrary uncalibrated cameras. In such a case, the scene may be reconstructed only up to a projective transformation of space. In the case where further assumptions are made about the cameras, it is possible to determine the scene more precisely. The most significant result in this areas is due to Maybank and Faugeras ([21]), who proved that if the calibration of the cameras is the same for all cameras, then three views are sufficient to determine the calibration of the camera, based only on point correspondences between the three images. Actually they proved only that three "motions" or pairs of views are sufficient, which is not quite the same thing. However, it has subsequently been established that three independent views (from which one may select three pairs) are sufficient ([20, 13]).

It is shown in [21] that each pair of views gives rise to two quadratic equations in the five unknown calibration parameters. These equations are known as Kruppa's equations ([18]).

If the calibration of the camera may be deduced from just three views, then the reconstruction problem is reduced to reconstruction using calibrated cameras. Several solutions of this problem (the relative orientation problem) have been proposed ([19, 16]). It is possible to reconstruct the scene up to a scaled Euclidean transformation, or similarity transform. It follows that Euclidean invariants may in theory be determined from three views of the scene with the same camera.

Unfortunately calibration and Euclidean reconstruction seems to be quite difficult to carry out in practice. Various algorithms have been proposed to do this ([20, 13]), but more work seems necessary. At present, the only way known of computing such Euclidean invariants is through reconstruction.

If further assumptions are made about the camera then self calibration becomes easier. For instance with a pair of (distinct) pinhole cameras (section 2) for which the principal point is also known it is possible to determine the focal lengths and carry out a scaled Euclidean reconstruction ([11]). Further results along this line are no doubt possible, since constraining the calibration simplifies Kruppa's equations.

6 Transfer

A number of papers have appeared in the last few years on the subject of *model transfer*. In this scenario, it is assumed that one has two (or more) images of an object or scene taken from two different viewpoints with uncalibrated cameras. These two views may be termed the *reference views*. Now, one is given another view of the scene in which it is possible to identify some of the points visible in the reference views. The task is to transfer the rest of the scene into the new view. In particular, one seeks to determine where the other points in the scene (appearing in the reference views) should appear in the new view.

As an example, suppose that a certain building is visible in the two reference views. In addition there are a number of terrain features or other reference points visible in the two reference views, which may also be located in the new view. By the process of transfer, one is able to overlay the image of the building on the new image, just as it ought to appear in the new image. This may be used to determine whether the building is present in the new view, or has been demolished. In this way, the two reference images serve instead of a complete geometric model of the building

– the need for a 3-dimensional model is replaced by two 2-dimensional views. Transfer is done without the need for camera calibration.

Epipolar Transfer Perhaps the first to put forward the idea of transfer was Eamon Barrett ([2]). His method though expressed in terms of the vanishing of a 9×9 determinant may be expressed in terms of the fundamental matrix. In particular, let J_0 and J_1 be the two reference views and let J_2 be the new view. Suppose that $\mathbf{u}_i^2$ are a set of at least 8 points in the view J_2 that may also be identified in the other images J_0 and J_1. Given 8 or more point correspondences between the images J_0 and J_2 one may compute the fundamental matrix F_{02}. Similarly one may compute the fundamental matrix F_{12}.

Now, suppose we are given another point that appears at coordinates $\mathbf{u}^0$ and $\mathbf{u}^1$ in the two reference images. Let $\mathbf{u}^2$ be the (unknown) coordinates of the point in the image J_2. By (4.1) $F_0 2\mathbf{u}^2$ represents the epipolar line corresponding to $\mathbf{u}^0$. In other words, point $\mathbf{u}^2$ lies on the line $F_{02}\mathbf{u}^0$. Similarly, $\mathbf{u}^2$ lies on the line $F_{12}\mathbf{u}^1$. One can therefore find $\mathbf{u}^2$ as the intersection of the two lines : $\mathbf{u}^2 = F_{12}\mathbf{u}^1 \times F_{02}\mathbf{u}^2$. This method of transfer is known as epipolar transfer.

Problems with Epipolar Transfer. The epipolar transfer method relies on finding the intersection of two epipolar lines. This method will be expected to fail if the two lines are close to being parallel, for then their exact intersection point may not be robustly determined. Let us consider when this will happen. Let $\mathbf{p}_{02}$ be the image of the centre of the view J_0 as seen in view J_2. Similarly $\mathbf{p}_{12}$ is the image of the centre of the view J_1 as seen in J_2. Points $\mathbf{p}_{02}$ and $\mathbf{p}_{12}$ are the two epipoles in image J_2. Now, let $\mathbf{u}^2$ be the coordinates of a point to be determined in image J_2. If this point lies in a straight line with the two epipoles $\mathbf{p}_{02}$ and $\mathbf{p}_{12}$ then it may not be determined by the intersection of two epipolar lines through the epipoles $\mathbf{p}_{02}$ and $\mathbf{p}_{12}$. However, $\mathbf{p}_{02}$, $\mathbf{p}_{12}$ $\mathbf{u}^2$ will lie in a straight line if the point $\mathbf{u}^2$ lies in the plane determined by the three camera centres. In short, the method of epipolar transfer will fail for points that lie close to the plane defined by the three camera centres. If the three camera centres lie in a straight line, then the method will fail entirely. This singularity is a deficiency of the method of epipolar transfer, rather than an intrinsic instability as we shall see.

Trilinear Relationship Amnon Shashua ([32]) has explored a different method of transfer that does not suffer from the deficiencies of the epipolar transfer method. Shashua shows that if $\mathbf{u}^0$, $\mathbf{u}^1$ and $\mathbf{u}^2$ are the coordinates of a point as seen in the three images, then there exist a pair of trilinear relations $f_1(\mathbf{u}^0, \mathbf{u}^1, \mathbf{u}^2) = 0$ and $f_2(\mathbf{u}^0, \mathbf{u}^1, \mathbf{u}^2) = 0$, where both f_1 and f_2 are trilinear functions in the homogeneous coordinates of $\mathbf{u}^0$, $\mathbf{u}^1$ and $\mathbf{u}^2$. The functions f_1 and f_2 are the same for all sets of points $\mathbf{u}^0$, $\mathbf{u}^1$ and $\mathbf{u}^2$, and are dependent only on the viewing parameters. Furthermore, the particular form of the functions f_1 and f_2 may be determined from only 7 point correspondences between all three images.

Now, given a point that is seen in positions $\mathbf{u}^0$ and $\mathbf{u}^1$ in the reference images, by substituting in the relations f_1 and f_2 one obtains a pair of linear equations $f_1(\mathbf{u}^2) = 0$ and $f_2(\mathbf{u}^2) = 0$. One may then solve these two equations to find $\mathbf{u}^2$. It may at first appear that once more there will exist degenerate conditions in which the two equations $f_1(\mathbf{u}^2) = 0$ and $f_2(\mathbf{u}^2) = 0$ do not have a unique solution. However, the particular form of the relations, described by Shashua, ensures that the set of points for which this occurs is a 1-dimensional set of points, rather than a 2-dimensional set (the plane defined by the camera centres) as in epipolar transfer.

Transfer by Projective Construction The method of Shashua requires 7 points in the new image to be located in each of the reference images. It will be shown here that in fact only 6 (actually $5\frac{1}{2}$) point matches are necessary to do transfer. In a typical situation in which the method of transfer is to be used, the two reference images are analyzed off-line. In general, a large number of points may be identified between the pair of reference images. Given sufficiently many points (at least 8) it is possible to do a complete projective reconstruction of the scene by any of the methods previously described in section 4.3. For greatest accuracy, an iterative method may be used to do the projective reconstruction. This is a one-time cost incurred during the building of the model. For instance, in the case of a building appearing in the scene, a complete projective model of the house may be built.

Now, suppose that a new image is given and a set of point matches $\mathbf{u}_i^0 \leftrightarrow \mathbf{u}_i^1 \leftrightarrow \mathbf{u}_i^2$ are found. From the correspondence $\mathbf{u}_i^1 \leftrightarrow \mathbf{u}_i^2$ the point location $\mathbf{x}_i$ in the 3D projective reconstruction may be computed, or is already known. Hence, we actually have a set of 3D - 2D correspondences $\mathbf{x}_i \mapsto \mathbf{u}_i^2$ between the point in the projective reconstruction and its image in the image J_2. Given $5\frac{1}{2}$ such correspondences, it is possible to determine a camera matrix P_2 such that $\mathbf{u}_i^2 = P\mathbf{x}_i$ for all i. This will be done by the method of direct linear transformation (DLT) as described by Sutherland ([33]). Now, any other point $\mathbf{x}$ in the model may be mapped to the point $P\mathbf{x}$ in the new image J_2.

Limitations of transfer. It is clear from the description of the reconstruction method, that any point known in the 3D projective model may be mapped uniquely into the new image J_2. The only way that the method may fail is when the position of a point may not be determined in the model. For instance, if $\mathbf{x}$ is a point lying on the line joining the camera centres of the two reference images, then its position may not be determined from its coordinates in the two reference images. This is so because the rays from the two camera centres to the point coincide, and hence their intersection is not well defined. Consequently, it is impossible to determine the location of the point in the 3D model, and hence in the new image J_2. This is an intrinsic limitation of transfer. It is impossible by any meanse to determine the location in a third image of a point lying on the line joining the centres of the two reference images (unless the third camera centre lies on this line as well). Thus, there exists a 1-dimensional critical set of points. This is a significant improvement over the situation with epipolar transfer where the set of critical points is complete plane of defined by the three camera centres.

Both the epipolar transfer and projective reconstruction methods generalize easily to the case where three or more reference views are given. If the camera centres of the reference views are non-collinear, then the critical set of points vanishes.

Further Reading

For projective geometry and homogeneous transformations Semple and Kneebone [29], Springer [6] and the appendix in [5]. A good discussion of affine transformations is given in Koenderink [17].

Part II
Single View Invariants

7 Invariants of planar objects

Algebraic invariants are invariants of configurations of algebraic objects, such as points, lines, conics, and cubics. The following lists a number of algebraic invariants to plane projective transformations.

7.1 Cross-ratio of four points on a line

This is a ratio of ratio of lengths on a line. It is preserved under projective transformations between lines

$$\begin{bmatrix} X_i' \\ 1 \end{bmatrix} = \begin{bmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \end{bmatrix} \begin{bmatrix} X_i \\ 1 \end{bmatrix} ***NOTATIONNEEDSFIXING*** \tag{12}$$

The cross-ratio is given by

$$I = \frac{(X_3' - X_1')(X_4' - X_2')}{(X_3' - X_2')(X_4' - X_1')} = \frac{(X_3 - X_1)(X_4 - X_2)}{(X_3 - X_2)(X_4 - X_1)} \tag{13}$$

where $\{X_1', X_2', X_3', X_4'\}$ represent the corresponding positions of each point along the line, e.g. $(X_3' - X_1')$ is the distance between points $\mathbf{X}_3'$ and $\mathbf{X}_1'$. Note, an invariant has the same value and *form* after the transformation.

Since points and lines are dual, there exists an equivalent cross-ratio for lines. The dual relation to collinearity is incidence at a point. A cross-ratio is defined on four lines which are incident at a single point.

7.2 Five points

Two functionally independent invariants can be constructed for five points $\mathbf{X}_i$ in the plane, no three of which are collinear:

$$I_1 = \frac{|\mathtt{S}_{431}||\mathtt{S}_{521}|}{|\mathtt{S}_{421}||\mathtt{S}_{531}|} \qquad I_2 = \frac{|\mathtt{S}_{421}||\mathtt{S}_{532}|}{|\mathtt{S}_{432}||\mathtt{S}_{521}|} \tag{14}$$

where $\mathtt{S}_{ijk} = (\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k)$ with $\mathbf{X}_i = (X_i, Y_i, 1)^\top$ and $|\mathtt{S}|$ the determinant of $\mathtt{S}$. The invariants are ratio of ratios of areas. The invariant can be derived by a geometric construction from the invariant of four concurrent lines, see Figure 4, and can also be derived algebraically using determinants (see below). The dual configuration is five general lines (no three concurrent). Examples of this invariant are given in figure 5 and table 2.

Proof of invariance

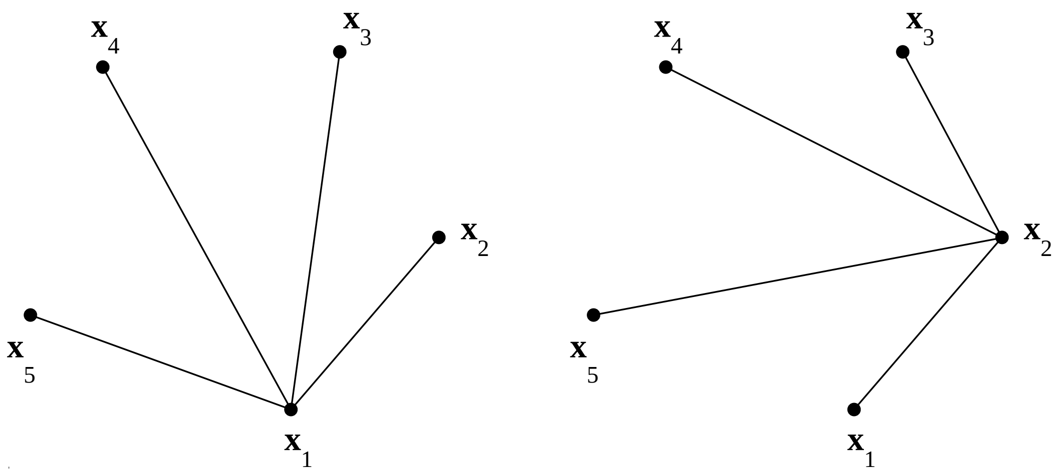


Figure 4: Five points on the plane have two functionally independent projective invariants. These can be expressed as the cross ratio of four lines from a base point to the other four points.

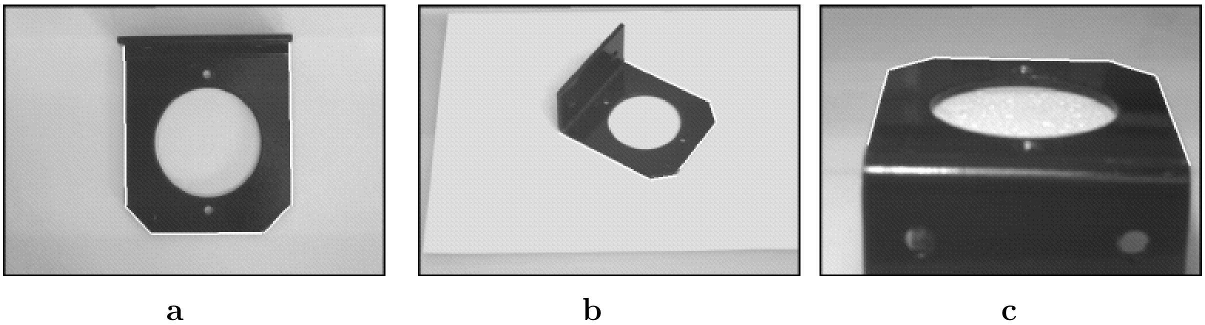


Figure 5: The lines used to compute the five line planar projective invariant for the above images are highlighted in white. The values are given in table 2.

Under the transformation $\mathbf{X}'_\mathbf{i} = k_i \mathtt{T}\mathbf{X}_\mathbf{i}$ (where $\mathbf{X'}_i = (X'_i, Y'_i, 1)^\top$ and $\mathbf{X}_i = (X_i, Y_i, 1)^\top$, and k_i accounts for scaling each point), a matrix of three such vectors transforms as:

$$\begin{bmatrix} X'_i & X'_j & X'_k \\ Y'_i & Y'_j & Y'_k \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{bmatrix} \begin{bmatrix} k_i X_i & k_j X_j & k_k X_k \\ k_i Y_i & k_j Y_j & k_k Y_k \\ k_i & k_j & k_k \end{bmatrix}$$

Taking determinants gives $|\mathtt{S'}_{ijk}| = k_i k_j k_k |\mathtt{T}||\mathtt{S}_{ijk}|$, where $\mathtt{S'}_{ijk} = (\mathbf{X'}_i, \mathbf{X'}_j, \mathbf{X'}_k)$. Dividing such expressions, as in equation (14) cancels both $|\mathtt{T}|$ and the scaling factors k_i giving, for example,

$$I_1 = \frac{|\mathtt{S'}_{431}||\mathtt{S'}_{521}|}{|\mathtt{S'}_{421}||\mathtt{S'}_{531}|} = \frac{|\mathtt{S}_{431}||\mathtt{S}_{521}|}{|\mathtt{S}_{421}||\mathtt{S}_{531}|} \quad \square$$

7.3 Conic and two points

An invariant can be formed from a conic, $\mathtt{C}$, and two points in general position, $\mathbf{X}_1$ and $\mathbf{X}_2$

$$I = \frac{(\mathbf{X}_1^\top \mathtt{C}\mathbf{X}_2)^2}{(\mathbf{X}_1^\top \mathtt{C}\mathbf{X}_1)(\mathbf{X}_2^\top \mathtt{C}\mathbf{X}_2)} \qquad (15)$$

Five line invariants		
Measured on	I_1	I_2
Object	0.840	1.236
Figure 5(a)	0.842	1.234
Figure 5(b)	0.840	1.232
Figure 5(c)	0.843	1.234

Conic and line pair invariant	
Measured on	I_1
Object	1.33
Figure 7(a)	1.33
Figure 7(b)	1.31
Figure 7(c)	1.28

Table 2: Values of plane projective invariants measured on the object, and from images with varying perspective effects. The values vary (due to measurement noise) by less than 0.4% for the five line invariants, and less than 4.0% for the conic and two line invariant.

Figure 6: (a) Two points, $\mathbf{X}_1$ and $\mathbf{X}_2$, coplanar with a conic $\mathtt{C}$ have a single plane projective invariant. This is related to the cross ratio of four collinear points, $\mathbf{X}_1, \mathbf{X}_i, \mathbf{X}_j$ and $\mathbf{X}_2$, (b), by a construction preserved by projection.

The invariant can be constructed geometrically from the cross ratio of points on a line, see figure 6. By duality the corresponding invariant for two lines, $\mathbf{L}_1$, $\mathbf{L}_2$, and a conic, $\mathbf{C}$, is

$$I = \frac{(\mathbf{L}_1^\top \mathtt{C}^{-1} \mathbf{L}_2)^2}{(\mathbf{L}_1^\top \mathtt{C}^{-1} \mathbf{L}_1)(\mathbf{L}_2^\top \mathtt{C}^{-1} \mathbf{L}_2)} \tag{16}$$

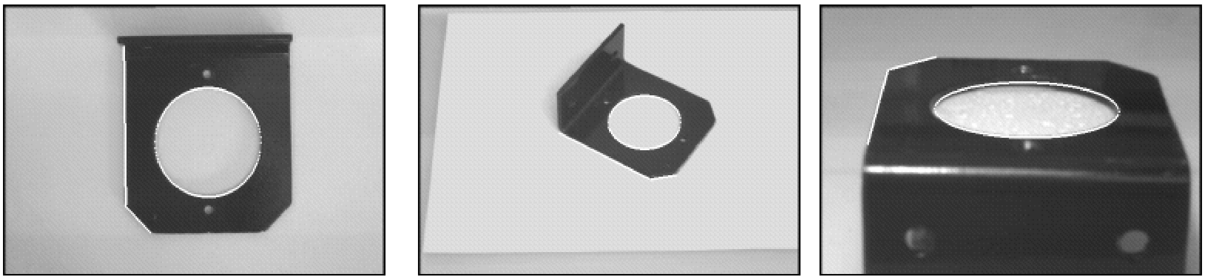
where the dual of a conic is a *line conic* given by $\mathtt{C}^{-1}$. Examples of this invariant are given in figure 7 and table 2.

7.4 Two conics

Two projectively invariant measures can be formed for a pair of conics. These are:

$$\begin{aligned} I_1 &= Trace[\mathtt{C}_1^{-1}\mathtt{C}_2](|\mathtt{C}_1|/|\mathtt{C}_2|)^{\frac{1}{3}} \\ I_2 &= Trace[\mathtt{C}_2^{-1}\mathtt{C}_1](|\mathtt{C}_2|/|\mathtt{C}_1|)^{\frac{1}{3}} \end{aligned}$$

where $\mathtt{C}_i$ is the conic matrix.



a b c

Figure 7: The curves used to compute the conic and two line planar projective invariant for the above images are highlighted in white. The values are given in table 2.

7.5 Canonical frame invariants

TO BE ADDED.

Further Reading

For algebraic invariants see Weiss [7] and the introduction to [5]. For canonical frame invariants see Lamdan *et al.* [3], and Rothwell *et al.* [26].

8 Application: Model based recognition

The task of model based recognition is to determine which, if any, of a set of known objects appear in a given image or image sequence. Object recognition systems draw on a library of geometric models, containing information about the shape and appearance of a set of known objects. Recognition is considered successful if the geometric configuration in an image can be explained as a perspective projection of a geometric model of the object.

Recognition, then, is the establishment of a correspondence between image and model features. Achieving this correspondence can be partitioned into three stages that should be contained within any recognition system:

Grouping: what subset of the data belongs to a single object?

Indexing: which object model projects to this data subset?

Verification: how much image support is there for this correspondence?

In the following we contrast a typical system which does not use invariants with one that does.

8.1 Transformation/pose based methods

Model hypotheses are generated exhaustively by matching image groups to all model features in the library. Each image-group to model-group match determines the transformation which projects the model into the image. The hypothesis is verified by comparing the projected model to image features. Essentially this method omits the indexing stage.

In this case, the complexity of constructing a recognition hypothesis is $O(\lambda i^k m^k)$ where λ is the number of models, i the number of image features, m the number of features per model, and k the number of features needed to determine the object-image transformation. In the SCERPO system developed by Lowe $k = 4$. Clearly, complexity increases linearly with the number of models in the library.

8.2 Index function methods

Model hypotheses are generated by an *index function* which directly identifies a model in the library from a set of grouped image features. Index functions must be *viewpoint invariant*, i.e. they are unaffected by the perspective distortions which occur when an object is projected into an image, naturally invariants play a key role here. The identified model is projected onto the image, and hypothesis verification applied.

In this case, the complexity of constructing a recognition hypothesis becomes $O(i^k)$ rather than $O(\lambda i^k m^k)$ in the pose based method, where k is the number of features required to form the index. So, unlike pose based recognition methods, recognition complexity need not be proportional to the number of models in the library. This is a considerable advantage if the number of models is large. Index functions are generally based on geometric invariants of the objects in the library.

8.3 Recognition Using Invariants

The two stages of model based vision using invariants are

1. **Model acquisition**
   Models are acquired directly from images. For planar objects this involves computing their plane projective invariants and storing their outline for the process of verification.

2. **Recognition**
   Invariants are computed for geometric configurations in the target image. If the invariant value corresponds to one in the model library a recognition hypothesis is generated. This is implemented by using invariant values to index into a hash table (no search). The hypothesis is confirmed or rejected by verification: The model outline from the acquisition image is projected onto the target image. If the projected edges overlap image edges to a sufficient extent then the hypothesis is *verified*.

Camera calibration is not required at any stage. An example is shown in figure 8 for an object (the bracket) which can be modelled by algebraic curves (namely 5 lines and a conic). It is recognised despite being partially occluded.

Further Reading

Recognition using projective invariants: [2, 27, 25].

9 Invariants of 3D objects

Much recent debate has focused around a theorem, proven by a number of authors [3, 1, 4], which states that invariants can not be measured for a 3D set of points in general position from a single

a **b**

Figure 8: (a) Acquisition image for bracket. Algebraic plane projective invariants are measured from this image. (b) The recognised bracket is highlighted in white.

view. The theorem has frequently been misinterpreted to mean that *no* invariants can be formed for three dimensional objects from a single image. For the theorem to hold, however, the points must be completely unconstrained, (like a cloud of gnats). If a 3D structure is *constrained*, then invariants are available.

3D projective invariants are invariant under projective transformations of $\mathcal{P}^3$. A projective transformation of $\mathcal{P}^3$can be written as:

$$\begin{bmatrix} X'_1 \\ X'_2 \\ X'_3 \\ X'_4 \end{bmatrix} = \begin{bmatrix} t_{11} & t_{12} & t_{13} & t_{14} \\ t_{21} & t_{22} & t_{23} & t_{24} \\ t_{31} & t_{32} & t_{33} & t_{34} \\ t_{41} & t_{42} & t_{43} & t_{44} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

with

$$x' = X'_1/X'_4 \qquad y' = X'_2/X'_4 \qquad z' = X'_3/X'_4$$

Fifteen parameters are required to define the 3D projective transformation matrix up to an arbitrary scale factor. Thus five 3D points are sufficient to construct a projective coordinate system. A sixth point will then have invariant 3D coordinates in the projective basis defined by the other five. These 3D point invariants can also be interpreted as the cross ratio of tetrahedral volumes define by taking determinants of point coordinates, four at a time.

For example, an invariant for six 3D points is given by

$$I_3(\mathbf{x}) = \frac{|\mathbf{x}_1\ \mathbf{x}_2\ \mathbf{x}_3\ \mathbf{x}_4|\,|\mathbf{x}_1\ \mathbf{x}_2\ \mathbf{x}_5\ \mathbf{x}_6|}{|\mathbf{x}_1\ \mathbf{x}_2\ \mathbf{x}_3\ \mathbf{x}_5|\,|\mathbf{x}_1\ \mathbf{x}_2\ \mathbf{x}_4\ \mathbf{x}_6|} \tag{17}$$

where $\mathbf{x}_i = (x_i, y_i, z_i, 1)^\top$. This invariant has the familiar property of invariants that

$$I_3(\mathbf{x}'_1, \mathbf{x}'_2, \mathbf{x}'_3, \mathbf{x}'_4, \mathbf{x}'_5, \mathbf{x}'_6) = I_3(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5, \mathbf{x}_6)$$

By assuming a set of constraints hold among the 3D projective invariants of a point set, it becomes possible to measure 3D projective invariants in a single view. The following section illustrates the nature of these constraints and provides a geometric interpretation for the measurable invariants.

9.1 Invariants

1. "Butterfly configuration"
   Points constrained to lie on two planes in a "butterfly" configuration, figure **??**, have a cross-ratio that can be measured in the image. This is a projective invariant of the entire 3D structure, and not simply a disguised planar invariant, since each plane contains only four points (five coplanar points are required to form a plane projective invariant from points alone).

2. *** MORE TO BE ADDED.

Further Reading

Polyhedra and symmetric point sets [28, 25].

Part III
Multiple View Invariants

10 Projective Invariants of Point Sets

As has been shown in section 4.3, although it is impossible to determine the exact geometry of a scene from multiple views, it is in general possible to reconstruct the scene up to an unknown projective transformation of space. Then projective invariants of the 3D structure (section 9) computed from a projective reconstruction of the scene will have the same value as if it were constructed from the actual scene. Such projective invariants do not include such scene properties as angles and length ratios, which are not invariant under projective transformations of the scene. The general strategy of computing these invariants is as follows.

1. Use image correspondences to compute the fundamental matrix F. Then find a factorization $F = [\mathbf{t}]_\times M$, and hence two camera matrices $P = [I \mid \mathbf{0}]$ and $P' = [M \mid \mathbf{t}]$.

2. Compute the projective reconstruction of the scene, using for instance the method of Section 4.3.

3. Compute a projective invariant of the reconstructed scene in $\mathcal{P}^3$.

10.1 An invariant of 6 points

We return to the six point invariant of equation (17). Given a set of six points $\{\mathbf{x}_i\}$ in $\mathcal{P}^3$, a coordinate system may be selected in which the first five points have coordinates $[1,0,0,0]^\top$, $[0,1,0,0]^\top$, $[0,0,1,0]^\top$, $[0,0,0,1]^\top$ and $[1,1,1,1]^\top$. The coordinates of the sixth point give rise to three independent projective invariants of the six points.

Another formulation of these invariants is given by selecting $\mathbf{x}_0$ and $\mathbf{x}_1$ as base points. Given any other point in $\mathcal{P}^3$, not collinear with $\mathbf{x}_0$ and $\mathbf{x}_1$, there exists a unique plane passing through that point and the two base points $\mathbf{x}_0$ and $\mathbf{x}_1$. In this way, the four points $\mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4$ and $\mathbf{x}_5$ give rise to four planes all containing the line joining $\mathbf{x}_0$ to $\mathbf{x}_1$. From the four planes it is possible to define a cross ratio. In particular, if λ is any line in space, skew to the line passing through $\mathbf{x}_0$ and $\mathbf{x}_1$, then λ intersects the four planes at points $\mathbf{p}_2$, $\mathbf{p}_3$, $\mathbf{p}_4$ and $\mathbf{p}_5$. The cross ratio of these four points on the line λ is a projective invariant of the six original points in $\mathcal{P}^3$. Different invariants result from different choices of $\mathbf{x}_0$ and $\mathbf{x}_1$.

Both these definitions of invariants fail if three of the points happen to be collinear, however, this case will be ignored for the sake of simplicity.

11 Projective Invariants of Lines

In this section, invariants of lines in space will be described. It will be shown that four lines in the 3-dimensional projective plane, $\mathcal{P}^3$give rise to two independent invariants under collineations of $\mathcal{P}^3$. Two different ways of defining invariants will be described, one algebraic and one geometric.

11.1 Computing Lines in Space

To be able to compute invariants of lines in space, it is necessary to be able to compute the location of a line in $\mathcal{P}^3$from its images in two or more views; Lines in the image plane are represented as 3-vectors. For instance, a vector $\boldsymbol{\lambda} = [\lambda, \mu, \nu]^\top$ represents the line in the plane given by the equation $\lambda u + \mu v + \nu w = 0$. Similarly, a plane in 3-dimensional space is represented in homogeneous coordinates as a 4-dimensional vector. The relationship between lines in the image space and corresponding planes in object space is given by the following lemma.

Lemma 11.10. *The set of all points in $\mathcal{P}^3$that are mapped by a camera with matrix P onto a line $\boldsymbol{\lambda}$ in the image is the plane π with coordinates $P^\top \boldsymbol{\lambda}$.*

Proof. A point $\mathbf{x}$ lies on π if and only if $P\mathbf{x}$ lies on the line $\boldsymbol{\lambda}$, and so $\boldsymbol{\lambda}^\top P\mathbf{x} = 0$. On the other hand, a point $\mathbf{x}$ lies on the plane π if and only if $\pi^\top \mathbf{x} = 0$. Comparing these two conditions leads to the conclusion that $\pi^\top = \boldsymbol{\lambda}^\top P$ or $\pi = P^\top \boldsymbol{\lambda}$ as required. □

If a line in space is seen in two or more views, then it may be found by computing the intersection of the corresponding planes in space.

11.2 Algebraic Invariant Formulation

Consider four lines λ_i in space. A line may be given by specifying either two points on the line or dually, two planes that meet in the line. It does not matter in which way the lines are described. For instance, in the formulae (19) and (20) below certain invariants of lines are defined in terms of pairs of points on each line. The same formulae could be used to define invariants in which lines are represented by specifying a pair of planes that meet along the line. Since the method of determining lines in space from two view given in section 11.1 gives a representation of the line as an intersection of two planes, the latter interpretation of the formulae is most useful.

Nevertheless, in the following description, of algebraic and geometric invariants of lines, lines will be represented by specifying two points, since this method seems to allow easier intuitive understanding. It should be borne in mind, however, that the dual approach could be taken with no change whatever to the algebra, or geometry.

In specifying lines, each of two points on the line will be given as a 4-tuple of homogeneous coordinates, and so each line λ_i is specified as a pair of 4-tuples

$$\lambda_i = ((a_{i1}, a_{i2}, a_{i3}, a_{i4})(b_{i1}, b_{i2}, b_{i3}, b_{i4}))$$

Now, given two lines λ_i and λ_j, one can form a 4×4 determinant, denoted by

$$|\lambda_i \lambda_j| = \det \begin{bmatrix} a_{i1} & a_{i2} & a_{i3} & a_{i4} \\ b_{i1} & b_{i2} & b_{i3} & b_{i4} \\ a_{j1} & a_{j2} & a_{j3} & a_{j4} \\ b_{j1} & b_{j2} & b_{j3} & b_{j4} \end{bmatrix} . \tag{18}$$

Finally, it is possible to define two independent invariants of the four lines by

$$I_1(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = \frac{|\lambda_1 \lambda_2| \ |\lambda_3 \lambda_4|}{|\lambda_1 \lambda_3| \ |\lambda_2 \lambda_4|} \tag{19}$$

and

$$I_2(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = \frac{|\lambda_1\lambda_2| \ |\lambda_3\lambda_4|}{|\lambda_1\lambda_4| \ |\lambda_2\lambda_3|} \ . \tag{20}$$

It is necessary to prove that the two quantities so defined are indeed invariants under collineations of $\mathcal{P}^3$. First, it must be demonstrated that the expressions do not depend on the specific formulation of the lines. That is, there are an infinite number of ways in which a line may be specified by designating two points lying on it, and it is necessary to demonstrate that choosing a different pair of points to specify a line does not change the value of the invariants. To this end, suppose that $[a_{i1}, a_{i2}, a_{i3}, a_{i4}]^\top$ and $[b_{i1}, b_{i2}, b_{i3}, b_{i4}]^\top$ are two distinct points lying on a line λ_i, and that $[a'_{i1}, a'_{i2}, a'_{i3}, a'_{i4}]^\top$ and $[b'_{i1}, b'_{i2}, b'_{i3}, b'_{i4}]^\top$ are another pair of points lying on the same line. Then, there exists a 2×2 matrix D_i such that

$$\begin{bmatrix} a'_{i1} & a'_{i2} & a'_{i3} & a'_{i4} \\ b'_{i1} & b'_{i2} & b'_{i3} & b'_{i4} \end{bmatrix} = D_i \begin{bmatrix} a_{i1} & a_{i2} & a_{i3} & a_{i4} \\ b_{i1} & b_{i2} & b_{i3} & b_{i4} \end{bmatrix} \ .$$

Consequently,

$$\begin{bmatrix} a_{i1} & a_{i2} & a_{i3} & a_{i4} \\ b_{i1} & b_{i2} & b_{i3} & b_{i4} \\ a_{j1} & a_{j2} & a_{j3} & a_{j4} \\ b_{j1} & b_{j2} & b_{j3} & b_{j4} \end{bmatrix} = \begin{bmatrix} D_i & 0 \\ 0 & D_j \end{bmatrix} \begin{bmatrix} a'_{i1} & a'_{i2} & a'_{i3} & a'_{i4} \\ b'_{i1} & b'_{i2} & b'_{i3} & b'_{i4} \\ a'_{j1} & a'_{j2} & a'_{j3} & a'_{j4} \\ b'_{j1} & b'_{j2} & b'_{j3} & b'_{j4} \end{bmatrix} \ .$$

Taking determinants, it is seen that the net result of choosing a different representation of the lines λ_i and λ_j is to multiply the value of $|\lambda_i\lambda_j|$ by a factor $\det(D_i)\det(D_j)$. Since each of the lines λ_i appears in both the numerator and denominator of the expressions (19) and (20), the factors will cancel and the values of the invariants will be unchanged.

Next, it is necessary to consider the effect of a change of projective coordinates. If H is a 4×4 invertible matrix representing a coordinate transformation of $\mathcal{P}^3$, then it may be applied to each of the points used to designate the four lines. The result of applying this transformation is to multiply the determinant $|\lambda_i\lambda_j|$ by a factor $\det(H)$. The factors on the top and bottom cancel, leaving the values of the invariants (19) and (20) unchanged. This completes the proof that I_1 and I_2 defined by (19) and (20) are indeed projective invariants of the set of four lines.

An alternative invariant may be defined by

$$I_3(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = \frac{|\lambda_1\lambda_4| \ |\lambda_2\lambda_3|}{|\lambda_1\lambda_3| \ |\lambda_2\lambda_4|} \ . \tag{21}$$

It is easily seen, that $I_3 = I_1/I_2$. However, if $|\lambda_1\lambda_2|$ vanishes, then both I_1 and I_2 are zero, but I_3 is in general non-zero. This means that I_3 can not always be deduced from I_1 and I_2. A preferable way of defining the invariants of four lines is as a homogeneous vector

$$I(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = [|\lambda_1\lambda_2| \ |\lambda_3\lambda_4| \, , \, |\lambda_1\lambda_3| \ |\lambda_2\lambda_4| \, , \, |\lambda_1\lambda_4| \ |\lambda_2\lambda_3|] \ . \tag{22}$$

Two such computed invariant values are deemed equal if they differ by a scalar factor. Note that this definition of the invariant avoids problems associated with vanishing or near-vanishing of the denominator in (19) or (20).

11.3 Degenerate Cases

The determinant $|\lambda_i \lambda_j|$ as given in (18) will vanish if and only if the four points involved are coplanar, that is, exactly when the two lines are coincident (meet in space). If all three components of the vector $I(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ given by (22) vanish, then the invariant is undefined. Enumeration of cases indicates that there are two essentially different configurations of lines in which this occurs.

1. Three of the lines lie in a plane.

2. One of the lines meets all the other three.

The configuration where one line meets two of the other lines is not degenerate, but does not lead to very much useful information, since two of the components of the vector vanish. Up to scale, the last component may be assumed to equal 1, which means that two such configurations can not be distinguished. In fact any two such configurations are equivalent under collineation.

11.4 Geometric Invariants of Lines

It is also possible to define projective invariants of sets of four lines geometrically. In particular, given four lines in space in general position, there will exist exactly two *transverse* lines that meet all four of these lines. The cross ratio of the points of intersection of lines with each of the transverse lines give two independent projective invariants of the set of four lines. These invariants may take real or complex values. The relationship of these invariants to the algebraic invariants is clarified in [12]. In particular, it is shown that there are just two independent projective invariants of four lines in space.

12 Geometric Approach to Invariants

An interesting approach to invariant computation has been developed by Gros and Quan ([7, 8, 9]) based on Geometric construction. The following description of this technique is somewhat different in detail from their approach, but the basic idea is the same.

The Coplanarity Test Gros and Quan make use of a test for coplanarity of four points, that I have first seen referred to in [5], ascribed to Roger Mohr. I describe here a different test for coplanarity, that seems to be very slightly simpler. Consider four points $\mathbf{u}_i \leftrightarrow \mathbf{u}'_i$ for $i = 1, \ldots, 4$ appearing in two images. Let $\mathbf{x}_i$ be the points in $\mathcal{P}^3$corresponding to these image points. We assume that the epipoles $\mathbf{p}$ and $\mathbf{p}'$ in the two images are also known. The epipole $\mathbf{p}$ is the point where the camera centre of the second camera appears in the first image, and $\mathbf{p}'$ is symmetrically defined. We assume that none of the points $\mathbf{u}_i$ or $\mathbf{u}'_i$ corresponds with one of the epipoles.

Proposition 12.11. *The four points $\mathbf{x}_i$ lie in a plane if and only if there is a 2D projective transformation taking $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4, \mathbf{p}$ to $\mathbf{u}'_1, \mathbf{u}'_2, \mathbf{u}'_3, \mathbf{u}'_4, \mathbf{p}'$.*

The fact that the sets of points are in projective equivalence means of course that the cross-ratio invariants of the two point sets are equal.

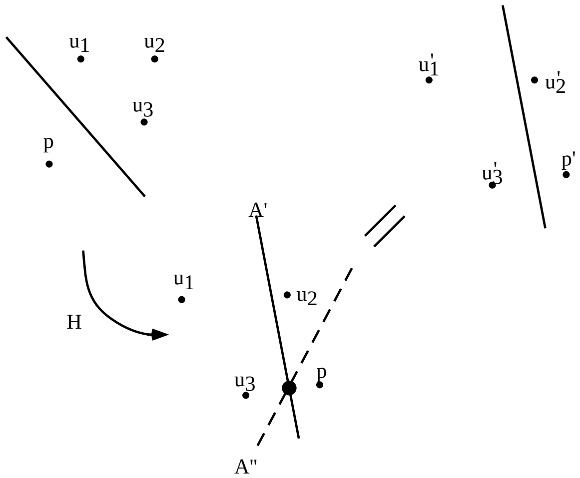


Figure 9: **Intersection of a line with a plane defined by three points.** *Consider points u_1, u_2, u_3 and the epipole p in one image and corresponding points u'_1, u'_2, u'_3 and p' in the second image. Compute the 2D projective mapping H that takes the points to their corresponding points in the second image. Transfer the line by H and compute its intersection with the line in the second image. This intersection point is the point where the line meets the plane of the three points, as seen in the second image. lines.*

Proof. Suppose that the four points $\mathbf{x}_i$ lie in a plane π and let $\mathbf{e}$ be the point where the line joining the two camera centres meets π. Then both sets of points $\mathbf{u}_1, \ldots, \mathbf{u}_4, \mathbf{p}$ and $\mathbf{u}'_1, \ldots, \mathbf{u}'_4, \mathbf{p}'$ are (2D) projectively equivalent to the set $\mathbf{x}_1, \ldots, \mathbf{x}_4, \mathbf{e}$, and hence to each other. To prove the converse, suppose that π is the plane containing $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$ and as before, let $\mathbf{e}$ be the intersection of π with the line joining the camera centres.

There exists a unique 2D projective transformations T taking $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{p}$ to $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{e}$, since a 2D transform is uniquely defined by 4 points. A transform T' may be similarly defined, and $T'^{-1}T$ is the unique transform taking $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{p}$ to $\mathbf{u}'_1, \mathbf{u}'_2, \mathbf{u}'_3, \mathbf{p}'$. If the point $\mathbf{x}_4$ does not lie on the plane π, however, then its projections onto π from the two camera centres are different. This means that $T\mathbf{x}_4 \neq T'\mathbf{x}_4$, and so $T'^{-1}T\mathbf{u}_4 \neq \mathbf{u}'_4$, so the points are not in projective correspondence. $\square$

Gros and Quan use a coplanarity criterion to allow them to compute the point of intersection of a line with a plane defined by three points. Using the above coplanarity criterion, this may be done as follows. Let $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ be the three points and let L be a line. Let the images of the points and line be $\mathbf{u}_i$ and $\boldsymbol{\lambda}$ in one image, and the same with primes in the other image. Let $\mathbf{p}$ and $\mathbf{p}'$ be the epipoles. Consider the transformation T defined by $T\mathbf{u}_i = \mathbf{u}'_i$ and $T\mathbf{p} = \mathbf{p}'$. The line $\boldsymbol{\lambda}'$ is also transformed by T to a line which we may (somewhat loosely) denote $T\boldsymbol{\lambda}$. The intersection of $\boldsymbol{\lambda}'$ and $T\boldsymbol{\lambda}$ is the image of the point where the L meets the plane π. This follows immediately from Proposition 12.11, since this is the unique point on $\boldsymbol{\lambda}'$ that is in projective correspondence, via T, with a point on the line $\boldsymbol{\lambda}$. This is illustrated in fig 12.

Three points and Two Lines. The idea put forward by Gros and Quan is to use this method to compute invariants. As an example, consider three points and two lines in 3D, and suppose

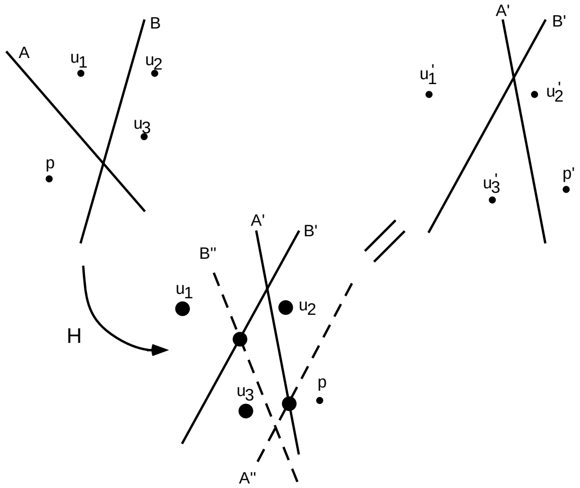


Figure 10: **Invariants from 3 points and 2 lines.** *At the top are two views of 3 points and 2 lines. It is assumed that the epipoles p and p' are also known. To compute the invariants of the set of points and lines, one finds a 2D projective transformation H that takes the points u_i to u_i' and p to p'. Since a 2D projectivity is determined by 4 points, the transform H is uniquely defined. Let $A \leftrightarrow A'$ and $B \leftrightarrow B'$ be the two matching lines. By transforming (warping) the first image by H, point u_i is mapped to u_i', and the lines A and B are mapped to lines A'' and B''. The points u_1', u_2', u_3' along with $u_4' = A' \cap A''$ and $u_5' = B' \cap B''$ form a set of 5 coplanar points (in the image plane of the primed camera). The two 2D projective invariants of these 5 points are invariants of the set of 3 points and 2 lines.*

that the epipoles are known. The three points define a plane. The intersection of the two lines with this plane, plus the three original points give five points in a plane, from which one may derive two invariants. By using the construction of the previous paragraph and applying it to both of the lines, one immediately finds the image of the three points and the two intersection points, as seen in the second (primed) image. This method is explained further in Fig 12.

Four Points and One Line. As another example, consider four points and one line. One may extract four subsets of three points from among the set of points. Each such subset defines a plane that meets the line in a single point. This provides four points on one line, and hence a single invariant. Using the above construction, one easily computes the four points, as seen in the primed image. This is illustrated in Fig 12.

Six Points The construction used for four points and one line can be used for six points by selecting two of the points to define a line. One is thus reduced to the case of four points and one line. The difference, however is that one now has two extra points on the line, namely the two points used to define the line. Thus we have a total of six points on the line A', namely the four intersection points plus the two points defining A'. From these six points on a line, we may extract three independent cross ratios.

The invariants described in the paragraphs above were computed by similar (but slightly different) constructions in [7, 8, 9].

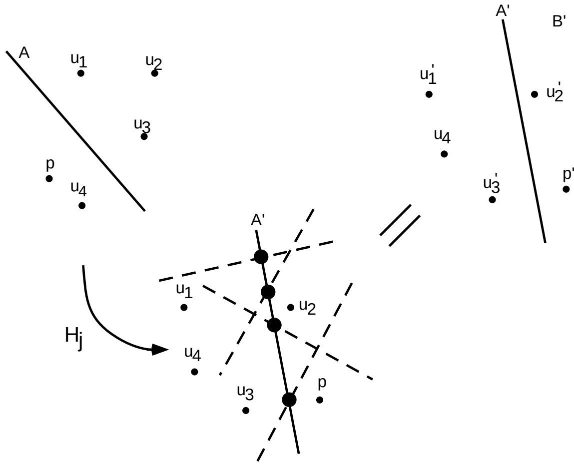


Figure 11: **Invariants from 4 points and 1 line.** *At the top are two views of 4 points and 1 lines. It is assumed that the epipoles p and p' are also known. For $j = 1, \ldots, 4$ one finds 2D projective transforms H_j defined as follows : H_j is the transform that takes p to p' and u_i to u'_i for each $i = 1, \ldots, 4$,* **except** *for $i = j$. By mapping the line A into the other image by transformations H_j one obtains four lines A''_j in the second image. The cross-ratio of the four points of intersection of A''_j with A' is an invariant of the four points and one line in space. This is because these four intersection points are in projective correspondence with the points of intersection of the line with the four planes defined by sets of three points.*

Further Reading

Other methods of geometric computation of invariants are given by Ponce ([23]) Quan ([24]) and in [14]. This includes invariants derived from smaller numbers of points (6 points in 3 views or 7 points in 2 views).

13 Algebraic Approach to Invariants

A very interesting algebraic method of computation of invariants was given by Carlsson ([4]). By completely algebraic techniques, involving the so-called *double algebra*, he derived explicit formulas for projective invariants of point and line sets as seen in a pair of images. The formulae express the invariants directly in terms of the image coordinates of the points as seen in the two views.

As an example, we consider the line invariants discussed in section 11. Thus, let $\boldsymbol{\ell}_i \leftrightarrow \boldsymbol{\ell}'_i$ for $i = 1, \ldots, 4$ be a set of four corresponding lines in two views, and let $\boldsymbol{\lambda}_i$ be the corresponding line in 3-space. Denote by $\mathbf{u}_{ij}$ the intersection of the lines $\boldsymbol{\ell}_i$ and $\boldsymbol{\ell}_j$. Thus $\mathbf{u}_{ij} = \boldsymbol{\ell}_i \times \boldsymbol{\ell}_j$. Similarly, let $\mathbf{u}'_{ij} = \boldsymbol{\ell}'_i \times \boldsymbol{\ell}'_j$. Then, with $|\boldsymbol{\lambda}_i \boldsymbol{\lambda}_j|$ defined as in (18), Carlsson shows that

$$|\boldsymbol{\lambda}_i \boldsymbol{\lambda}_j| = k \mathbf{u}'^{\top}_{ij} F \mathbf{u}_{ij} \tag{23}$$

where k is a constant. From this it follows that the invariant (22) may be written entirely in terms of the image coordinates, and the fundamental matrix.

$$I = (\mathbf{u}'^{\top}_{12} F \mathbf{u}_{12} . \mathbf{u}'^{\top}_{34} F \mathbf{u}_{34},\ \mathbf{u}'^{\top}_{13} F \mathbf{u}_{13} . \mathbf{u}'^{\top}_{24} F \mathbf{u}_{24},\ \mathbf{u}'^{\top}_{14} F \mathbf{u}_{14} . \mathbf{u}'^{\top}_{23} F \mathbf{u}_{23}) \ .$$

The formula (23) may be seen directly using as follows. The determinant $|\boldsymbol{\lambda}_i\boldsymbol{\lambda}_j|$ may be defined directly by expressing the line $\boldsymbol{\lambda}_i$ as the intersection of the two planes $P^\top\ell_i$ and $P'^\top\ell'_i$. From this one directly obtains

$$|\boldsymbol{\lambda}_i\boldsymbol{\lambda}_j| = \det\left[P^\top\ell_i, P'^\top\ell'_i, P^\top\ell_j, P'^\top\ell'_j\right] \enspace .$$

However, one may verify that

$$\det\left[P^\top\ell_i, P'^\top\ell'_i, P^\top\ell_j, P'^\top\ell'_j\right] = [\boldsymbol{\ell}'_i \times \boldsymbol{\ell}'_j]^\top F[\boldsymbol{\ell}_i \times \boldsymbol{\ell}_j] \enspace .$$

This may be verified first for elementary basis vectors such as $[1,0,0]^\top$ and then extended by linearity to arbitary $\boldsymbol{\ell}_i$ and $\boldsymbol{\ell}'_i$. This derivation was suggested to me by conversations with Rajiv Gupta.

By consideration of the results of Carlsson, it may be seen that many of the invariants of mixed point and line sets in $\mathcal{P}^3$are actually just disguised invariants of sets of lines. In fact, replacing two points by the line that passes through them and then defining a line invariant gives the same result. This is certainly true of all the invariants considered in [4].

Invariants of Points Given a set of six points, $x_1, \ldots, x_6$, denote by $\boldsymbol{\lambda}_{ij}$ the line that passes through points $\mathbf{x}_i$ and $\mathbf{x}_j$. Then the expression

$$I_1(\lambda_{12}, \lambda_{45}, \lambda_{13}, \lambda_{46}) = \frac{|\lambda_{12}\lambda_{45}| \; |\lambda_{13}\lambda_{46}|}{|\lambda_{12}\lambda_{46}| \; |\lambda_{13}\lambda_{45}|} \tag{24}$$

is invariant. The lines $\boldsymbol{\lambda}_{ij}$ map to lines $\boldsymbol{\ell}_{ij}$ and $\boldsymbol{\ell}'_{ij}$ in the two images, and these lines may be computed easily in terms of the measured image coordinates $\mathbf{u}_i$ and $\mathbf{u}'_i$. Then, applying (23), one easily obtains a formula for the six-point invariant in terms of the image coordinates $\mathbf{u}_i$ and $\mathbf{u}'_i$ and the fundamental matrix F.

$$I = (\mathbf{x}_1, \ldots, \mathbf{x}_6) = \frac{\mathbf{u}'_{12:45}{}^\top F\mathbf{u}_{12:45}.\mathbf{u}'_{13:46}{}^\top F\mathbf{u}_{13:46}}{\mathbf{u}'_{12:46}{}^\top F\mathbf{u}_{12:46}.\mathbf{u}'_{13:45}{}^\top F\mathbf{u}_{13:45}} \tag{25}$$

where $\mathbf{u}_{ij:kl}$ is the intersection of the line $\mathbf{u}_i \times \mathbf{u}_j$ with the line $\mathbf{u}_k \times \mathbf{u}_l$. Namely, $\mathbf{u}_{ij:kl} = (\mathbf{u}_i \times \mathbf{u}_j) \times (\mathbf{u}_k \times \mathbf{u}_l)$. This formula is formula (30) in [4]. The lines $\boldsymbol{\lambda}_{12}$ and $\boldsymbol{\lambda}_{13}$ meet in one point, $\mathbf{x}_1$, and similarly lines $\boldsymbol{\lambda}_{45}$ and $\boldsymbol{\lambda}_{46}$ meet in $\mathbf{x}_4$. Consequently, there is only one invariant associated with the set of four lines. By choosing other combinations of the points to construct lines, one obtains more invariants. In total, there are three independent invariants of a set of six points.

Invariants of Mixed Points and Lines Given a set of mixed points and lines in $\mathcal{P}^3$, one can define invariants in a similar manner by joining subsets of the points together to make lines. Then, one applies the 4-line invariant to obtain invariants of the set of points and lines.

References

[1] K.E. Atkinson. *An Introduction to Numerical Analysis, 2nd Edition.* John Wiley and Sons, New York, 1989.

[2] Eamon. B. Barrett, Michael H. Brill, Nils N. Haag, and Paul M. Payton. Invariant linear methods in photogrammetry and model matching. In J.L. Mundy and A. Zisserman, editors, *Geometric Invariance in Computer Vision*, pages 277 – 292. MIT Press, Boston, MA, 1992.

[3] J. B. Burns, R. S. Weiss, and E. M. Riseman. The non-existence of general-case view-invariants. In J. Mundy and A. Zisserman, editors, *Geometric Invariance in Computer Vision.* MIT Press, 1992.

[4] Stefan Carlsson. Multiple image invariants using the double algebra. In *Proc. of the Second Europe-US Workshop on Invariance, Ponta Delgada, Azores*, pages 335–350, October 1993.

[5] O. D. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In *Computer Vision - ECCV '92, LNCS-Series Vol. 588, Springer-Verlag*, pages 563 – 578, 1992.

[6] O. D. Faugeras, Q.-T Luong, and S. J. Maybank. Camera self-calibration: Theory and experiments. In *Computer Vision - ECCV '92, LNCS-Series Vol. 588, Springer-Verlag*, pages 321 – 334, 1992.

[7] P. Gros and L. Quan. Projective Invariants for Vision. Technical Report RT 90 IMAG - 15 LIFIA, Irimag–Lifia, Grenoble, France, December 1992.

[8] P. Gros and L. Quan. 3D Projective Invariants from Two Images. In *Geometric Methods in Computer Vision II, SPIE 1993 International Symposium on Optical Instrumentation and Applied Science*, pages 75–86, July 1993.

[9] Patrick Gros. 3D projective invariants from two images. In *Proc. of the Second Europe-US Workshop on Invariance, Ponta Delgada, Azores*, pages 65–85, October 1993.

[10] R. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 761–764, 1992.

[11] R. I. Hartley. Estimation of relative camera positions for uncalibrated cameras. In *Computer Vision - ECCV '92, LNCS-Series Vol. 588, Springer-Verlag*, pages 579 – 587, 1992.

[12] R. I. Hartley. Camera calibration using line correspondences. In *Proc. DARPA Image Understanding Workshop*, pages 361–366, 1993.

[13] Richard I. Hartley. Euclidean reconstruction from uncalibrated views. In *Proc. of the Second Europe-US Workshop on Invariance, Ponta Delgada, Azores*, pages 187–202, October 1993.

[14] Richard I. Hartley. Projective reconstruction and invariants from multiple images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 16:1036–1041, October 1994.

[15] Richard I. Hartley. Projective reconstruction from line correspondences. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 903–907, 1994.

[16] B. K. P. Horn. Relative orientation. *International Journal of Computer Vision*, 4:59 – 78, 1990.

[17] J. J. Koenderink. Optic flow. *Vision Research*, 26(1):161–179, 1986.

[18] E. Kruppa. Zur Ermittlung eines Objektes aus zwei Perspektiven mit innerer Orientierung. *Sitz.-Ber. Akad. Wiss., Wien, math. naturw. Abt. IIa.*, 122:1939–1948, 1913.

[19] H.C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, Sept 1981.

[20] Q.-T Luong. *Matrice Fondamentale et Calibration Visuelle sur l'Environnement*. PhD thesis, Universite de Paris-Sud, Centre D'Orsay, 1992.

[21] 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.

[22] R. Mohr, L. Quan, F. Veillon, and B. Boufama. Relative 3D reconstruction using multiples uncalibrated images. Technical Report RT 84-I-IMAG LIFIA 12, Irimag–Lifia, 1992.

[23] Jean Ponce, Todd Cass, and David Marimont. Relative stereo and motion reconstruction. Report UIUC-BI-AI-RCV-93-07, Beckman Institute, University of Illinois, 1993.

[24] L. Quan. Invariants of 6 Points from 3 Uncalibrated Images. Rapport Technique RT 101 IMAG 19 LIFIA, LIFIA-IMAG, Grenoble, October 1993. To appear in ECCV94.

[25] C. Rothwell, D. Forsyth, A. Zisserman, and J. Mundy. Extracting projective information from single views of 3D point sets. Technical Report 1973/93, Oxford University Department of Engineering Science, 1993.

[26] C. Rothwell, A. Zisserman, D. Forsyth, and J. Mundy. Canonical frames for planar object recognition. LNCS 588, 1992.

[27] C. Rothwell, A. Zisserman, C. I. Marinos, D. Forsyth, and J. Mundy. Relative motion and pose from arbitrary plane curves. *Image and Vision Computing*, 10(4):250–262, 1992.

[28] C. A. Rothwell. *Recognition Using Projective Invariance*. PhD thesis, 1993.

[29] J.G. Semple and G. T. Kneebone. *Algebraic Projective Geometry*. Oxford University Press, Oxford, 1952.

[30] Amnon Shashua. On geometric and algebraic aspects of 3D affine and projective structures from perspective 2D views. In *Applications of Invariance in Computer Vision : Proc. of the Second Joint European - US Workshop, Ponta Delgada, Azores – LNCS-Series Vol. 825, Springer Verlag*, pages 127 – 144, October 1993.

[31] Amnon Shashua. Projective depth: A geometric invariant for 3D reconstruction from two perspective/orthographic views and for visual recognition. In *Proc. International Conference on Computer Vision*, pages 583–590, 1993.

[32] Amnon Shashua. Algebraic functions for recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 17(8):779–789, August 1995.

[33] I.E. Sutherland. Sketchpad: A man-machine graphical communications system. Technical Report 296, MIT Lincoln Laboratories, 1963. Also published by Garland Publishing Inc, New York, 1980.

References

[1] Clemens, D.T. and Jacobs, D.W. "Model Group Indexing for Recognition," Proceedings CVPR91, p.4-9, 1991, and *IEEE Trans. PAMI,* Vol. 13, No. 10, p.1007-1017, October 1991.

[2] Forsyth, D.A., Mundy, J.L., Zisserman, A.P., Coelho, C., Heller, A. and Rothwell, C.A., Invariant Descriptors for 3-D Object Recognition and Pose, *PAMI*-13, No. 10, p.971-991, October 1991.

[3] Lamdan, Y., Schwartz, J.T. and Wolfson, H.J., Object Recognition by Affine Invariant Matching, *Proc. CVPR,* p.335-344, 1988.

[4] Moses, Y. and Ullman, S. "Limitations of Non Model-Based Recognition Systems," Proceedings ECCV2, p.820-828, 1992.

[5] Mundy, J. L. and Zisserman, A. (editors), "Geometric Invariance in Computer Vision,", MIT Press, Cambridge Ma, 1992.

[6] Springer, C.E., *Geometry and Analysis of Projective Spaces,* Freeman, 1964.

[7] Weiss, I., Projective Invariants of Shapes, *Proc. CVPR,* p.291-297, 1988.