The Geometry of Two Views

Roberto Cipolla
Department of Engineering, University of Cambridge

Extracted from Chapter 5, Visual Motion of Curves and Surfaces by R. Cipolla and P.J. Giblin, Cambridge University Press 1999

1 Camera model for perspective projection onto image plane

To reconstruct 3D space from images we require the mapping from image plane pixel coordinates, $\mathbf{u}$, to visual rays in a fixed *world* coordinate system, $\mathbf{p}$. This is determined by transformations describing the position and orientation of the coordinate system attached to the camera relative to the world coordinate system; perspective (central) projection onto the image plane and the geometry of the CCD array. These three transformations are derived below and can be conveniently written as a 3×4 projection matrix (Roberts 1965).

Property 1.1 The projection matrix. *Under perspective projection the map between the three-dimensional world coordinates of a point* (X, Y, Z) *and its two-dimensional image plane pixel coordinates* (u, v) *can be written as a linear mapping in homogeneous coordinates and represented by a* 3×4 *projection matrix:*

$$\begin{bmatrix} \zeta u \\ \zeta v \\ \zeta \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}. \tag{1}$$

Rigid-body transformation

Consider a coordinate system $\mathbf{X} = (X, Y, Z)$ attached to the world reference frame, and another coordinate system $\mathbf{X}_c = (X_c, Y_c, Z_c)$ attached to the camera at position $\mathbf{c}(t)$, where the optical axis is aligned with Z_c. See Figure 1.

The camera and reference coordinate systems are related by a rigid body transformation which are conveniently represented with a rotation matrix and a translation vector, $\mathbf{t}$, by:

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} t_X \\ t_Y \\ t_Z \end{bmatrix} \tag{2}$$

where the translation vector is related to the position of the camera centre by

$$\mathbf{t} = -\mathbf{R}\mathbf{c}.$$

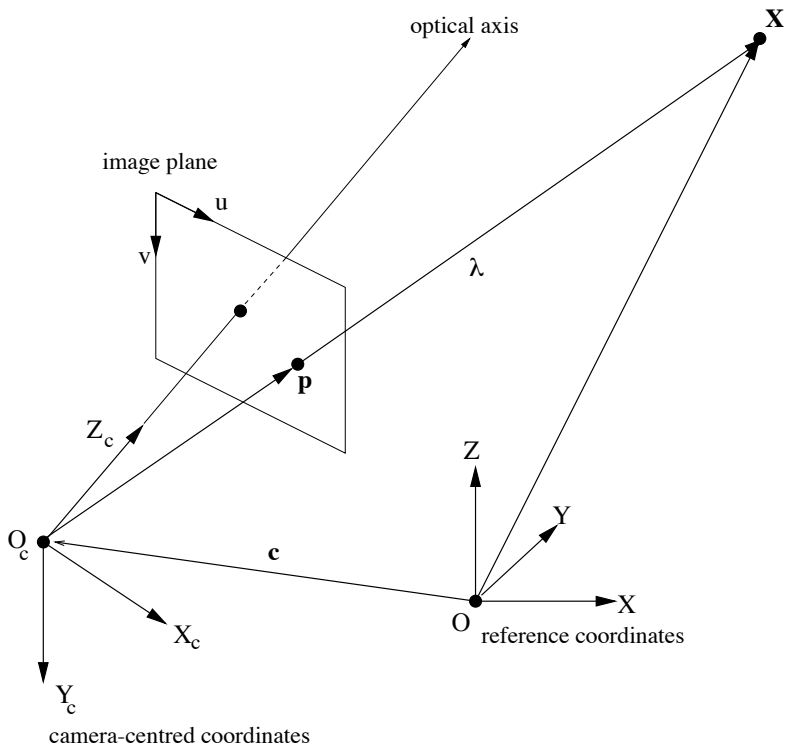


Figure 1: Camera model and camera and reference coordinate systems. $\mathbf{X} = \mathbf{c} + \lambda\mathbf{p}$ and $\mathbf{X} = \mathbf{c} + \mathbf{R}\mathbf{X}_c$.

Perspective projection onto the CCD image plane

Perspective projection onto the imaging plane followed by the conversion of image plane coordinates into CCD pixel coordinates, (u, v), can be modelled by

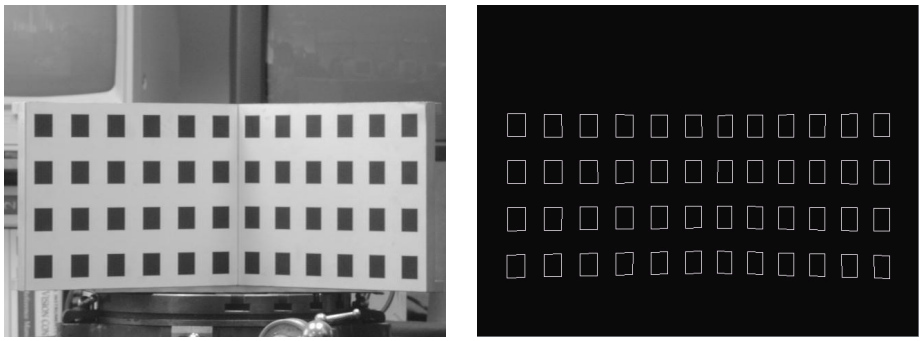
$$u = u_0 + \alpha_u \frac{X_c}{Z_c} \tag{3}$$

$$v = v_0 + \alpha_v \frac{Y_c}{Z_c} \tag{4}$$

where the CCD array axes are assumed aligned with the X_c and Y_c axes; (u_0, v_0) is the *principal point* (the point of intersection of the optical axis and the image plane); α_u and α_v are image scaling factors. These four parameters are known as the *internal camera parameters*. The ratio α_v/α_u is known as the aspect ratio.

Projection matrix

The relationship between image pixel coordinates and rays in Euclidean 3-space can now be expressed succinctly by introducing *homogeneous coordinates* to represent image points with 3-vectors and points in 3-space by 4-vectors, defined up to arbitrary scales (e.g. ζ). Homogeneous (projective) coordinates



(a) (b)

Figure 2: Camera calibration. A camera is calibrated by processing an image of a calibration grid (a). The image positions of known 3D points on the grid are extracted automatically. Edge detection is followed by fitting lines to the image segments. Intersections of lines are used to localize the image features to sub-pixel accuracy (b).

are often used in projective geometry and allow us to represent projective transformations as a matrix multiplications. By concatenating the matrices for the transformations described above the relationship becomes:

$$\begin{bmatrix} \zeta u \\ \zeta v \\ \zeta \end{bmatrix} = \begin{bmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_X \\ r_{21} & r_{22} & r_{23} & t_Y \\ r_{31} & r_{32} & r_{33} & t_Z \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

or more simply as the 3×4 *projection matrix* representing the perspective projection of a point in space onto a digitized image given in Property 1.1 where equality is defined up to an arbitrary scale.

$$\mathbf{u} = \mathbf{PX} \tag{5}$$

The projection matrix, $\mathbf{P}$, is not a general 3×4 matrix. It has 11 parameters (since the overall scale does not matter) and it can be decomposed into a 3×3 upper triangular matrix of camera internal parameters called the *camera calibration matrix*, $\mathbf{K}$, and a matrix representing the rigid-body motion.

The mapping from an image point to a visual ray in 3-space is expressed in homogeneous coordinates and up to an arbitrary scale by:

$$\mathbf{u} = \mathbf{KRp} \tag{6}$$

2 Camera model for weak perspective and orthographic projection

A useful approximation to perspective projection occurs when the field of view is narrow or the depth variation along the line of sight is small compared with

the distance from the camera to the scene. The camera model can then be simplified to *weak perspective* and (3) and (4) can be re-written as:

$$u = u_0 + \alpha_u \frac{X_c}{Z_o} \tag{7}$$

$$v = v_0 + \alpha_v \frac{Y_c}{Z_o}. \tag{8}$$

The difference with perspective projection is that all image points are scaled uniformly by Z_o, the mean distance of the features of the scene to the camera centre. Weak perspective can in fact be considered as the orthographic (parallel) projection of all points onto the plane $Z_c = Z_0$ followed by a perspective projection to give a uniform inverse-depth scaling. It can be represented by the transformation:

$$\begin{bmatrix} u_a \\ v_a \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_X \\ r_{21} & r_{22} & r_{23} & t_Y \\ 0 & 0 & 0 & Z_o \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

If the principal point of a camera is (u_0, v_0), the variation of depth in the scene is ΔZ along the optical axis and the mean distance of the features of the scene to the camera is Z_0, the difference of the image of a point taken from a perspective camera (u, v) and its image with the weak perspective camera, (u_a, v_a) is given by

$$u - u_a = (u - u_0)\Delta Z / Z_o \tag{9}$$

$$v - v_a = (v - v_0)\Delta Z / Z_o. \tag{10}$$

When the field of view is narrow, the terms $u - u_0$ and $v - v_0$ will be small. In this case, or when the depth variation of the scene is much smaller than its mean depth, e. g. $\Delta Z / Z_0 < 0.1$, the error due to the weak perspective appproximation is negligible.

Orthographic projection can be modelled in exactly the same way but with no scaling due to depth by setting $Z_c = f$.

3 Camera calibration

For reconstruction we require the camera centres, $\mathbf{c}(t)$, and the rays $\mathbf{p}(s,t)$. These can be obtained from the projection matrix for each viewpoint. *Camera calibration* is the name given to the process of recovering the projection matrix from an image of a controlled scene. For example, we might set up the camera to view the calibrated grid shown in Figure 2(a) and automatically extract the image positions of known 3D points (Figure 2(b)). Each image point, (u_i, v_i), of a known calibration point, X_i, Y_i, and Z_i, generates two equation which the elements of the projection matrix must satisfy:

$$u_i = \frac{\zeta u_i}{\zeta} = \frac{p_{11}X_i + p_{12}Y_i + p_{13}Z_i + p_{14}}{p_{31}X_i + p_{32}Y_i + p_{33}Z_i + p_{34}}$$

$$v_i = \frac{\zeta v_i}{\zeta} = \frac{p_{21}X_i + p_{22}Y_i + p_{23}Z_i + p_{24}}{p_{31}X_i + p_{32}Y_i + p_{33}Z_i + p_{34}}.$$

These equations can be rearranged to give two linear equations in the 12 unknown elements of the projection matrix. For n calibration points and their corresponding image projections we have $2n$ equations:

$$\begin{bmatrix} X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & -u_1X_1 & -u_1Y_1 & -u_1Z_1 & -u_1 \\ 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -v_1X_1 & -v_1Y_1 & -v_1Z_1 & -v_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_n & Y_n & Z_n & 1 & 0 & 0 & 0 & 0 & -u_nX_n & -u_nY_n & -u_nZ_n & -u_n \\ 0 & 0 & 0 & 0 & X_n & Y_n & Z_n & 1 & -v_nX_n & -v_nY_n & -v_nZ_n & -v_n \end{bmatrix} \begin{bmatrix} p_{11} \\ p_{12} \\ p_{13} \\ p_{14} \\ p_{21} \\ p_{22} \\ p_{23} \\ p_{24} \\ p_{31} \\ p_{32} \\ p_{33} \\ p_{34} \end{bmatrix} = \mathbf{0}.$$

Since there are 11 unknowns (scale is arbitrary), we need to observe at least 6 reference points to recover the projection matrix and calibrate the camera.

Numerical considerations

The equations can be solved using orthogonal least squares. First, we write the equations in matrix form:

$$\mathbf{Ax} = \mathbf{0} \tag{11}$$

where $\mathbf{x}$ is the 12×1 vector of unknowns (the 12 elements of the projection matrix, p_{ij}), $\mathbf{A}$ is the $2n \times 12$ matrix of measurements and n is the number of observed calibration points. A linear solution (least squares) which minimizes $||\mathbf{Ax}||$ subject to $||\mathbf{x}|| = 1$ is obtained as the unit eigenvector corresponding to the smallest eigenvalue of $\mathbf{A}^\top\mathbf{A}$. Numerically this computation is performed via the *singular value decomposition* of the matrix (Strang 1988)

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^\top$$

where $\mathbf{\Lambda} = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_{12})$ is the diagonal matrix of singular values and the matrices $\mathbf{U}$ and $\mathbf{V}$ are orthonormal. The columns of $\mathbf{V}$ are the eigenvectors of $\mathbf{A}^\top\mathbf{A}$ and the least squares solution is given by the last column of $\mathbf{V}$ which is the singular vector with the smallest singular value σ_{12}. The least squares solution is, however, only approximate and should be used as the starting point for non-linear optimization: i.e. finding the parameters of the projection matrix, $\mathbf{P}$, that minimize the errors between measured image points, (u_i, v_i) and the projections onto the image plane of the reference points:

$$\min_{\mathbf{P}} \sum_i ||(u_i, v_i) - \mathbf{P}(X_i, Y_i, Z_i, 1)||^2$$

Once the projection matrix has been estimated the first 3×3 submatrix, $(\mathbf{KR})^\top$, can be easily decomposed by standard matrix algorithms into an upper triangular matrix, $\mathbf{K}$, and a rotation (orthonormal) matrix (known as QR decompositon) or used directly to determine the ray in space $\mathbf{p}$ and the position of the camera centre:

$$\mathbf{p} = (\mathbf{KR})^{-1}\mathbf{u} \tag{12}$$

$$\mathbf{c} = -(\mathbf{KR})^{-1}(p_{14}, p_{24}, p_{34})^\top. \tag{13}$$

4 Epipolar geometry

Epipolar geometry plays a key part in the algorithms to recover structure and motion. We briefly review the geometry of two views and describe how to compute the epipolar geometry when the cameras are calibrated. The use of uncalibrated cameras and the recovery of the epipolar geometry from apparent contours is then described.

The epipolar constraint

In stereo vision the projection of a world point in two calibrated viewpoints can be used to recover the three-dimensional position by triangulation. The geometry of the two views, as shown in Figure 3, plays a key part in helping to find correspondences by constraining the search for correspondence from a region to a line. This matching constraint is known as the *epipolar constraint.*

The epipolar constraint arises from the fact that the two rays, $\mathbf{p}$ and $\mathbf{p}'$, to a common scene point, $\mathbf{X}$, and the optical centres of the two camera (the stereo baseline, $\mathbf{t} = \Delta\mathbf{c}$) lie in a plane called the *epipolar plane.* The intersection of the epipolar plane with each image plane defines a line called an *epipolar line.* The correspondence of an image point in the first view, $\mathbf{u}$, must line on the epipolar line, $\mathbf{l}'$, in the other view shown in Figure 3. Using homogeneous coordinates to represent the coefficients of a line in the image as a 3-vector, the epipolar constraints in each view can be written as:

$$\mathbf{u} \cdot \mathbf{l} = 0 \tag{14}$$

$$\mathbf{u}' \cdot \mathbf{l}' = 0. \tag{15}$$

Each world point, $\mathbf{X}$, has its own epipolar plane. The family of epipolar planes define a *pencil* of epipolar lines which pass through a common point called the *epipole*, illustrated in Figure 4. The epipoles and pencil of epipolar lines in each view are known as the *epipolar geometry.* The epipolar geometry is completely determined by the relative position, $\mathbf{t}$, and relative orientation, $\mathbf{R}$, of the two views and the camera parameters of each camera, $\mathbf{K}$ and $\mathbf{K}'$ respectively. It does not depend on the 3D structure of the scene being viewed.

The essential matrix

The epipolar constraint is a co-planarity constraint and can be expressed algebraically as a scalar triple product:

$$\mathbf{p}' \cdot (\mathbf{t} \wedge \mathbf{p}) = 0. \tag{16}$$

With out loss of generality, we can align the reference coordinate system with the second camera so that the epipolar constraint can be rewritten in terms of image positions (3-vectors in homogeneous coordinates), $\mathbf{u}$ and $\mathbf{u}'$:

$$\mathbf{u}'^{\top}\mathbf{K}'^{-\top}\mathbf{E}\mathbf{K}^{-1}\mathbf{u} = 0 \tag{17}$$

where $\mathbf{E}$ is a 3×3 matrix known as the *essential matrix* (Longuet-Higgins 1981) and is the product of a skew-symmetric or antisymmetric matrix (representing

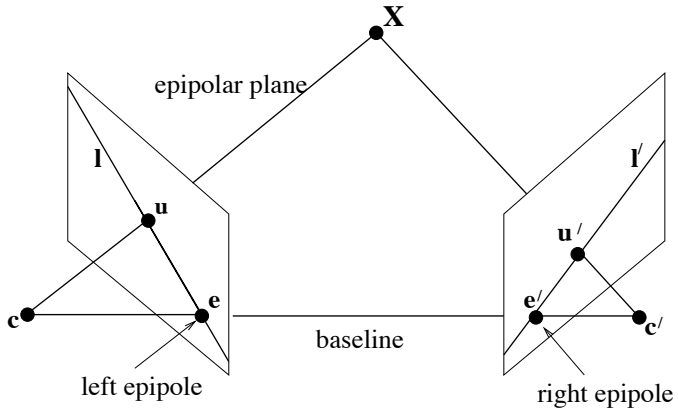


Figure 3: The geometry of two views. In stereo vision an *epipolar plane* is the plane defined by a 3D point $\mathbf{X}$ and the optical centres of the two cameras. The *baseline* is the line joining the optical centres. An *epipole* is the point of intersection of the baseline with the image plane. An *epipolar line*, $\mathbf{l}$ and $\mathbf{l}'$, is a line of intersection of the epipolar plane with an image plane. It is the image in one camera of the ray from the other camera's optical centre to the point $\mathbf{X}$.

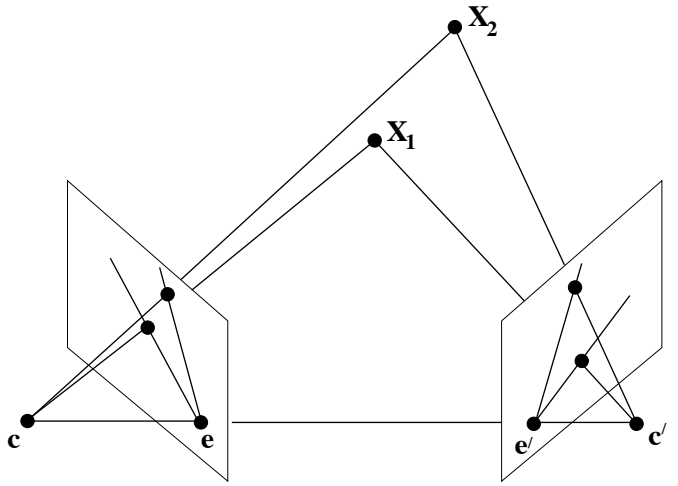


Figure 4: Epipolar geometry. Each world point $\mathbf{X}$ has its own epipolar plane which rotates about the baseline. All epipolar lines intersect at the epipole.

the vector product with the translation vector) and an orthonormal matrix representing the rotation between the two views:

$$\mathbf{E} = \mathbf{t} \wedge \mathbf{R} = [\mathbf{t}]_\times \mathbf{R}$$

where

$$[\mathbf{t}]_\times = \begin{bmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{bmatrix}$$

and $\mathbf{R}$ now specifies the relative orientation between the views.

The essential matrix is of maximum rank 2. Its factorization into the product of a non-zero skew-symmetric matrix and a rotation matrix is only possible if it has two equal non-zero singular values. The other is of course equal to zero (Tsai and Huang 1984, Faugeras and Maybank 1990).

The fundamental matrix

From (17) we see that the epipolar geometry can be conveniently specified by introducing a matrix, $\mathbf{F}$ (Faugeras 1992)

$$\mathbf{F} = \mathbf{K}'^{-\top} \mathbf{E} \mathbf{K}^{-1}. \tag{18}$$

Property 4.1 The epipolar constraint and the fundamental matrix. *The image coordinates (projective representation using homogeneous coordinates) of all pairs of corresponding points,* $\mathbf{u}_i = (u_i, v_i, 1)^\top$ *and* $\mathbf{u}'_i = (u'_i, v'_i, 1)^\top$*, must satisfy the epipolar constraint:*

$$\mathbf{u}'^\top_i \mathbf{F} \mathbf{u}_i = 0 \tag{19}$$

or

$$\begin{bmatrix} u'_i & v'_i & 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = 0 \tag{20}$$

where $\mathbf{F}$ *is a* 3×3 *real matrix of rank 2 which is defined up to an arbitrary scale and is known as the fundamental matrix.*

Epipolar lines and epipoles

The epipolar geometry (see Figure 4) is completely determined by the fundamental matrix.

Property 4.2 Epipolar geometry from fundamental matrix.

1. **Epipolar lines.**

   The epipolar line (represented by a homogeneous 3-vector), $\mathbf{l}'$*, corresponding to a point* $\mathbf{u}$ *in the other view is given by*

$$\mathbf{l}' = \mathbf{F}\mathbf{u} \tag{21}$$

and $\mathbf{u}'$ *must lie on this line to satisfy the epipolar constraint:*

$$\mathbf{u}' \cdot \mathbf{l}' = 0.$$

The epipolar line corresponding to $\mathbf{u}'$ *is given by* $\mathbf{l} = \mathbf{F}^\top \mathbf{u}'$.

2. **Epipoles.**

   The epipole is defined as the point in each image which is common to all the epipolar lines. The left and right epipoles ($\mathbf{e}$ *and* $\mathbf{e}'$ *in homogeneous coordinates) are therefore given by the null spaces of* $\mathbf{F}$ *and* $\mathbf{F}^\top$ *respectively*

$$\mathbf{F}\mathbf{e} = \mathbf{0} \tag{22}$$

$$\mathbf{F}^\top \mathbf{e}' = \mathbf{0}. \tag{23}$$

5 Epipolar geometry from projection matrices

For calibrated cameras with known projection matrices it is trivial to compute the fundamental matrix and hence obtain the epipolar geometry (epipoles and epipolar lines for each image feature). We here outline a simple method by exploiting the following result.

Property 5.1 Projective ambiguity. *The pair of cameras and projection matrices* $\mathbf{P}$ *and* $\mathbf{P}'$ *give rise to the same fundamental matrix as the pair of cameras and projection matrices* $\mathbf{PH}$ *and* $\mathbf{P}'\mathbf{H}$ *where* $\mathbf{H}$ *is a* 4×4 *non-singular matrix.*

A simple proof can be found in (Hartley 1992 and 1994) but follows trivially from the fact that the simultaneous transformation of the projection matrices, $\mathbf{P}$ by $\mathbf{H}$ and the 3D point coordinates, $\mathbf{X}$, by $\mathbf{H}^{-1}$ leaves the image coordinates $\mathbf{u} = \mathbf{PX}$, unchanged.

Assume we are given the projection matrices for two viewpoints, $\mathbf{P}$ and $\mathbf{P}'$. The position of the optical centre of the first camera, $\mathbf{c}$, can be computed directly from the projection matrix $\mathbf{P}$.

In homogeneous coordinates we can represent it by a 4-vector $\mathbf{C} = (\mathbf{c}^\top 1)$ so that

$$\mathbf{PC} = \mathbf{0}$$

and its projection into the second image plane defines the epipole, $\mathbf{e}'$,

$$\mathbf{e}' = \mathbf{P}'\mathbf{C}. \tag{24}$$

We can also compute the pseudo-inverse, $\mathbf{P}^+$, of the projection matrix $\mathbf{P}$,

$$\mathbf{P}^+ = \mathbf{P}^\top (\mathbf{P}\mathbf{P}^\top)^{-1}, \tag{25}$$

such that multiplication with the first projection matrix gives the identity matrix, $\mathbf{I}$, and multiplication with the second projection matrix gives a 3×3 matrix (a two-dimensional projective transformation), $\mathbf{M}$:

$$\mathbf{I} = \mathbf{P}\,\mathbf{P}^+ \tag{26}$$

$$\mathbf{M} = \mathbf{P}'\mathbf{P}^+ \tag{27}$$

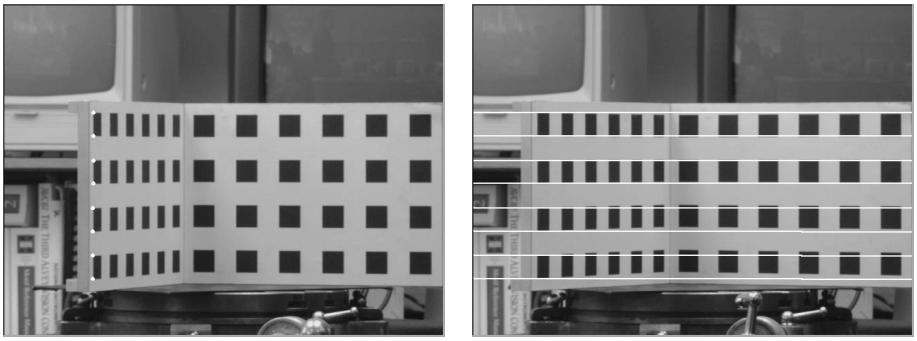


Figure 5: Epipolar geometry computed from known projection matrices. Selected image points are shown in the left view with corresponding epipolar lines shown in the right view. The corresponding image feature satisfies the epipolar constraint.

The two projection matrices have in this way been normalized to have the special forms:

$$\mathbf{PH} = [\mathbf{I} \mid \mathbf{0}] \tag{28}$$
$$\mathbf{P}'\mathbf{H} = [\mathbf{M} \mid \mathbf{e}'] \tag{29}$$

These normalized projection matrices are known as *canonical cameras*. From Property 5.1 both the original projection matrices and these normalized forms have the same fundamental matrix $\mathbf{F}$ given by (see Property 8.1):

$$\mathbf{F} = [\mathbf{e}']_\times \mathbf{M}. \tag{30}$$

An example of the epipolar geometry of two discrete views computed from calibrated projection matrices using (24) - (27) and (30) is shown in Figure 5.

For uncalibrated cameras we do not have the projection matrices and hence $\mathbf{E}$, $\mathbf{K}$ and $\mathbf{K}'$ are unknown *a priori*. The fundamental matrix and epipolar geometry, however, can still be estimated from *point* or curve correspondences between the two views.

The recovery of the structure and motion from point correspondences has attracted considerable attention and many practical algorithms exist to recover both the spatial configuration of the points and the viewer motion compatible with the views. These are briefly reviewed in §6 to §8.

6 The fundamental matrix from point correspondences

For uncalibrated cameras we do not have the projection matrices and hence the essential matrix parameters, $\mathbf{t}$ and $\mathbf{R}$, and the camera calibration parameters, $\mathbf{K}$ and $\mathbf{K}'$, are unknown a priori. The fundamental matrix, however, can be estimated from *point* correspondences between the two views. We briefly describe the algorithms used for computing the epipolar geometry from point correspondences (also reviewed in Zhang 1998) before describing the extension to curves and apparent contours.

From the epipolar constraint (20) we see that each point correspondence, $\mathbf{u}_i = (u_i, v_i, 1)^\top$ and $\mathbf{u'}_i = (u'_i, v'_i, 1)^\top$, generates one constraint on the epipolar geometry which can be expressed in terms of the elements of the fundamental matrix $\mathbf{F}$:

$$\begin{bmatrix} u'_i & v'_i & 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = 0.$$

For n pairs of correspondences, the constraints can be rearranged as linear equations in the 9 unknown elements of the fundamental matrix:

$$\begin{bmatrix} u'_1u_1 & u'_1v_1 & u'_1 & v'_1u_1 & v'_1v_1 & v'_1 & u_1 & v_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u'_nu_n & u'_nv_n & u'_1 & v'_nu_n & v'_nv_n & v'_n & u_n & v_n & 1 \end{bmatrix} \begin{bmatrix} f_{11} \\ f_{12} \\ f_{13} \\ f_{21} \\ f_{22} \\ f_{23} \\ f_{31} \\ f_{32} \\ f_{33} \end{bmatrix} = \mathbf{0}$$

or in matrix form:

$$\mathbf{Af} = \mathbf{0}$$

where $\mathbf{A}$ is an $n \times 9$ measurement matrix, and $\mathbf{f}$ represents the elements of the fundamental matrix as a 9-vector. Given 8 or more correspondences a solution[1] can be found by least squares as the unit eigenvector ($\mathbf{f}$ is defined up to an arbitrary scale) corresponding to the minimum eigenvalue of $\mathbf{A}^\top\mathbf{A}$. A unique solution is obtained unless the points and the camera centres lie on a ruled quadric or all the points lie on a plane (Faugeras and Maybank 1990).

The computation can be poorly conditioned and it is important to precondition the image points by normalizing them to improve the condition number of $\mathbf{A}^\top\mathbf{A}$ before estimating the elements of the fundamental matrix by singular value decomposition (Hartley 1998).

Parametrization of the fundamental matrix

Two steps can be taken to improve the solution. The most important requires enforcing the rank 2 property of the fundamental matrix. This can be achieved by a suitable parametrization of $\mathbf{F}$.

The epipolar geometry between two uncalibrated views is completely determined by 7 independent parameters: the position of the epipoles in the two views, $\mathbf{e} = (u_e, v_e, 1)^\top$ and $\mathbf{e'} = (u'_e, v'_e, 1)^\top$, and the 3 parameters of the one-dimensional projective transformation[2] relating the pencil of epipolar lines in

[1]Note that the fundamental matrix has only 7 degrees of freedom since its determinant must be zero. A non-unique solution can be obtained from only 7 point correspondences and is described in (Huang and Netravali 1994).

[2]This one-dimensional projective transformation (also known as a collineation or homography) can be represented by a 2×2 matrix in homogeneous coordinates.

view 1 to those in view 2 (Luong and Faugeras 1996),

$$\tau_i' = -\frac{h_2\tau_i + h_1}{h_4\tau_i + h_3} \tag{31}$$

where τ_i and τ_i' represent the directions (as the gradient of a line) of a pair of corresponding epipolar lines, $\mathbf{l}_i$ and $\mathbf{l}_i'$, in the first and second images respectively. Namely:

$$\tau_i = \frac{v_i - v_e}{u_i - u_e} \tag{32}$$

$$\tau_i' = \frac{v_i' - v_e'}{u_i' - u_e'}. \tag{33}$$

The transformation of epipolar lines between views is sometimes known as the *epipolar transformation* and is fixed by 3 pairs of epipolar line correspondences. The correspondence of any additional epipolar line is completely determined since it must preserve the cross-ratio of the 4 epipolar planes and corresponding epipolar lines. See (Luong and Faugeras 1996) and Figure 6.

Substituting (32) and (33) into (31) for the image coordinates of a pair of corresponding points results in the epipolar constraint and leads to the following minimal parametrization of the fundamental matrix:

$$\mathbf{F} = \begin{bmatrix} h_1 & h_2 & -u_e h_1 - v_e h_2 \\ h_3 & h_4 & -u_e h_3 - v_e h_4 \\ -u_e' h_1 - v_e' h_3 & -u_e' h_2 - v_e' h_4 & u_e u_e' h_1 + v_e u_e' h_2 + u_e v_e' h_3 + v_e v_e' h_4 \end{bmatrix} \tag{34}$$

This parametrization will be exploited later when apparent contours are used instead of point correspondences to estimate the epipolar geometry.

Optimization

Another improvement requires finding the 7 independent parameters of the fundamental matrix which minimize the distances between the image points and their epipolar lines.

Property 6.1 Geometric error using epipolar distances. *The geometric distance between an image point* $\mathbf{u}'$ *and the epipolar line,* $\mathbf{l}' = \mathbf{F}\mathbf{u}$ *is given by:*

$$\frac{(\mathbf{u}_i'^{\top}\mathbf{F}\mathbf{u}_i)^2}{(\mathbf{F}\mathbf{u}_i)_1^2 + (\mathbf{F}\mathbf{u}_i)_2^2} \tag{35}$$

A suitable cost function, C, consisting of the sum of the squared geometric distances (defined above) between image points and their epipolar lines in both images (Luong and Faugeras 1996),

$$C = \sum_i \left(\frac{1}{(\mathbf{F}\mathbf{u}_i)_1^2 + (\mathbf{F}\mathbf{u}_i)_2^2} + \frac{1}{(\mathbf{F}^{\top}\mathbf{u}'_i)_1^2 + (\mathbf{F}^{\top}\mathbf{u}'_i)_2^2} \right) (\mathbf{u}_i'^{\top}\mathbf{F}\mathbf{u}_i)^2$$

can be minimized by non-linear optimization techniques (Press et al. 1988).

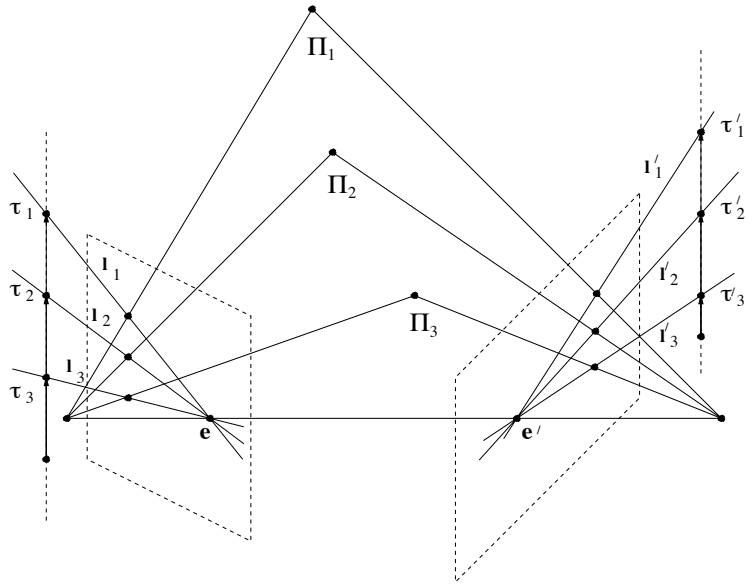


Figure 6: The epipolar geometry of an uncalibrated stereo pair of images is completely specified by the image positions of the epipoles and 3 pairs of corresponding epipolar lines. The projective parameters τ and τ' represent the intersection of the epipolar line and the line at infinity. The directions in the 2 views are related by a one-dimensional projective transformation (homography).

7 Recovery of the projection matrices and viewer motion

As shown above it is possible to recover the epipolar geometry (via the fundamental matrix) from point correspondences in the case of uncalibrated cameras. Nevertheless we must recover the projection matrices corresponding to each viewpoint if we are to attempt reconstruction.

Factorization of the essential matrix

If the camera internal parameters, $\mathbf{K}$ and $\mathbf{K}'$, are known the viewer motion and the projection matrices are determined by the epipolar geometry. We can transform the recovered fundamental matrix into an essential matrix (18):

$$\mathbf{E} = \mathbf{K}'^{\top}\mathbf{F}\mathbf{K} \tag{36}$$

and decompose this matrix into a skew-symmetric matrix corresponding to translation and an orthonormal matrix corresponding to the rotation between the views:

$$\mathbf{E} = [\mathbf{t}]_{\times}\mathbf{R}. \tag{37}$$

The latter is in fact only possible if the the essential matrix has rank 2 and two equal singular values (Tsai and Huang 1984). This property turns out to be very important in recovering constraints on the internal parameters of the cameras when they are uncalibrated. The difference in the two singular values can be used to refine the camera parameters. In fact, each fundamental matrix places two quadratic constraints on the internal calibration parameters (the Kruppa equations) which can be used to estimate, for example, the scale factors of the two cameras. This is known as self-calibration (Maybank and Faugeras 1992 and Hartley 1992).

The translation vector, $\mathbf{t}$, which can only be recovered up to an unknown magnitude, can be found as the unit eigenvector corresponding to the smallest eigenvalue of $\mathbf{E}\mathbf{E}^{\top}$ since it must satisfy

$$\mathbf{E}^{\top}\mathbf{t} = 0.$$

The rotation can then be obtained as the orthonormal matrix which minimizes the matrix Frobenius norm

$$||\mathbf{E} - [\mathbf{t}_{\times}]\mathbf{R}||^2$$

which can be solved linearly if we represent the rotation with a quaternion (Horn 1987).

Numerical considerations

An alternative numerical approach is to perform the singular value decomposition (Strang 1988) of the essential matrix (Hartley 1992):

$$\mathbf{E} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^{\top} \tag{38}$$

where $\mathbf{\Lambda} = \mathrm{diag}(\sigma_1, \sigma_2, \sigma_3)$ and the matrices $\mathbf{U}$ and $\mathbf{V}$ are orthogonal. The decomposition into a translation vector and the rotation between the two views requires that $\sigma_1 = \sigma_2 \neq 0$ and $\sigma_3 = 0$. The nearest (in the sense of minimizing the Frobenius norm between the two matrices) essential matrix with the correct properties can be obtained by setting the two largest singular values to be equal to their average and the smallest one to zero (Hartley 1992). The translation and axis and angle of rotation can then be obtained directly up to arbitrary signs and unknown scale for the translation:

$$[\mathbf{t}]_\times = \mathbf{U} \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{U}^\top \tag{39}$$

$$\mathbf{R} = \mathbf{U} \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{V}^\top \tag{40}$$

The projection matrices follow directly from the recovered translation and rotation by aligning the reference coordinate system with the first camera to give:

$$\begin{aligned} \mathbf{P} &= \mathbf{K}[\mathbf{I} \mid \mathbf{0}] \\ \mathbf{P}' &= \mathbf{K}'[\mathbf{R} \mid \mathbf{t}]. \end{aligned}$$

Four solutions are possible due to the arbitrary choice of signs for translation, $\pm\mathbf{t}$, and rotation, $\mathbf{R}$ or $\mathbf{R}^\top$. The correct solution is easily disambiguated by ensuring that reconstructed points lie in front of the cameras.

8 Recovery of the projection matrices for uncalibrated cameras

If the camera calibration matrices are unknown the projection matrices can not be uniquely recovered from the epipolar geometry of two views alone. In fact we will see that they can only be recovered up to an arbitrary 3D projective transformation, known as a projective ambiguity.

From (18) it follows that the fundamental matrix can, like the essential matrix, be factorized into a skew-symmetric matrix corresponding to translation and a 3×3 non-singular matrix (ignoring arbitrary scalings of the elements of $\mathbf{F}$):

$$\begin{aligned} \mathbf{F} &= \mathbf{K}'^{-\top}[\mathbf{t}]_\times \mathbf{R}\mathbf{K}^{-1} \\ &= [\mathbf{K}'\mathbf{t}]_\times \mathbf{K}'\mathbf{R}\mathbf{K}^{-1} \\ &= [\mathbf{e}']_\times \mathbf{M}_\infty \end{aligned} \tag{41}$$

where

$$\mathbf{M}_\infty = \mathbf{K}'\mathbf{R}\mathbf{K}^{-1} \tag{42}$$

is a 2D projective transformation (homography) which maps points on the plane at infinity in one image to the other (Luong and Vieville 1996).

The factorization of the fundamental matrix as the product of a skew-symmetric matrix and a non-singular matrix $\mathbf{M}$ is not unique since there is a 3 parameter family of matrices $\mathbf{M}$ (which represents the 2D projective transformation between views induced by different planes) such that:

$$\mathbf{M} = \mathbf{M}_\infty + \mathbf{e}'\mathbf{v}^\top$$

where $\mathbf{v}$ can be arbitrarily chosen to give a different projective transformation but the same fundamental matrix (Hartley 1994 and Luong and Vieville 1996)[3].

Property 8.1 Factorization of the fundamental matrix. *The fundamental matrix can be factorized into a skew-symmetric matrix and a* 3×3 *non-singular matrix,* $\mathbf{M}$*:*

$$\mathbf{F} = [\mathbf{e}']_\times \mathbf{M}$$

where $\mathbf{e}'$ *is equivalent to the epipole in the second view:*

$$\mathbf{F}^\top \mathbf{e}' = \mathbf{0}$$

and $\mathbf{M}$ *can be chosen from the 3-parameter family (defined by the arbitrarily choice of* $\mathbf{v}$*) of homographies given by:*

$$\mathbf{M} = [\mathbf{e}']_\times \mathbf{F} + \mathbf{e}'\mathbf{v}^\top.$$

Singular value decomposition of the fundamental matrix

As with the essential matrix, we can factorize the fundamental matrix into a skew-symmetric component and a non-singular matrix by analysing its singular value decomposition:

$$\mathbf{F} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^\top$$

where $\mathbf{\Lambda} = \mathrm{diag}(r, s, 0)$. The skew-symmetric component can be recovered from:

$$[\mathbf{e}']_\times = \mathbf{U} \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{U}^\top \tag{43}$$

in exactly the same way as with calibrated cameras. The non-singular matrix $\mathbf{M}$ is no longer an orthogonal transformation and is not uniquely defined. As shown by Property 8.1, the homography (two-dimensional projective transformation) is defined up to an arbitrary choice of parameters, here described by $\{\alpha, \beta, \gamma\}$:

$$\mathbf{M} = \mathbf{U} \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r & 0 & 0 \\ 0 & s & 0 \\ \alpha & \beta & \gamma \end{bmatrix} \mathbf{V}^\top. \tag{44}$$

[3]Note the relationship with the minimal parametrization introduced in (34). A 3-parameter family of two-dimensional projective transformations, $\mathbf{M}$, representing the transformation between views induced by points on a plane, can be recovered from the pair of epipoles and 3 pairs of corresponding epipolar lines. The epipoles must satisfy $\mathbf{e}' = \mathbf{M}\mathbf{e}$ while the epipolar lines must pass through the epipoles and satisfy $\mathbf{l}'_i = \mathbf{M}^{-\top}\mathbf{l}_i$.

Canonical cameras and projective ambiguity

The factorization of the fundamental matrix can be used to compute the canonical cameras – the normalized projection matrices – given by (28) and (29)

$$\begin{aligned} \mathbf{PH} &= [\mathbf{I} \mid \mathbf{0}] \\ \mathbf{P}'\mathbf{H} &= [\mathbf{M} \mid \mathbf{e}'] \end{aligned}$$

The real projection matrices, $\mathbf{P}$ and $\mathbf{P}'$, have only been recovered up to an arbitrary 3D projective transformation represented algebraically by a 4×4 matrix $\mathbf{H}$, and known as a projective ambiguity.

Property 8.2 Projective ambiguity. *A general 3D projective transformation can be represented by a non-singular* 4×4 *matrix,* $\mathbf{H}$*, of the form*

$$\mathbf{H} = \begin{bmatrix} s\mathbf{R}_\mathbf{w} & \mathbf{t}_\mathbf{w} \\ \mathbf{0}^\top & 1 \end{bmatrix} \begin{bmatrix} \mathbf{K}^{-1} & \mathbf{0} \\ \mathbf{0}^\top & 1 \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{v}^\top & 1 \end{bmatrix}. \tag{45}$$

The projective ambiguity is composed of the following effects. A metric transformation resulting from the rigid body motion between the coordinate system of the first camera and the reference frame and an arbitrary scaling. This can be ignored if we align the reference coordinate system with the first camera and accept that shape can only be recovered up to an arbitrary scale, s, if the distance between the two camera centres is unknown. The second component of the ambiguity results from an 3D affine transformation due to the unknown parameters of the first camera. Finally we are left with a projective transformation which transforms points on the plane $(\mathbf{v}^\top 1)\mathbf{X} = 0$ to points on the plane at infinity and results from the ambiguity of Property 8.2.

The ambiguity in the projection matrices is of the form above and will result in a projective ambiguity in the recovered geometry, i.e. the 3D coordinates of visible points, $\mathbf{X}$, can only be recovered up to a 3D projective transformation, $\mathbf{H}^{-1}\mathbf{X}$. This ambiguity can only be removed with additional information derived from scene constraints or knowledge of the camera parameters, $\mathbf{K}$ and $\mathbf{K}'$. In particular the ambiguity is completely removed by using the 3D position of 5 known scene points to determine the transformation $\mathbf{H}$ or $\mathbf{H}^{-1}$. Alternatively we require the internal camera parameters of the first camera and must then find the equation of the plane at infinity represented by $\mathbf{v}$ where

$$\mathbf{M} = \mathbf{K}'\mathbf{R}\mathbf{K}^{-1} + \mathbf{e}'\mathbf{v}^\top.$$

The ambiguity is removed by using knowledge of the camera parameters to fix $\mathbf{v}$ to make the homography, $\mathbf{M}$, be that induced by the plane at infinity (Hartley 1994). The rotation matrix follows.