SECOND EDITION

Multiple View Geometry in computer vision

Richard Hartley and Andrew Zisserman

CAMBRIDGE

CAMBRIDGE more information - www.cambridge.org/9780521540513

Multiple View Geometry in Computer Vision

Second Edition

Richard Hartley
Australian National University,
Canberra, Australia

Andrew Zisserman
University of Oxford, UK

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo, Delhi, Dubai, Tokyo, Mexico City

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521540513

First published 2000
Reprinted 2001, 2002
Second Edition 2003
9th printing 2011

Printed in the United Kingdom at the University Press, Cambridge

A catalogue record for this publication is available from the British Library

ISBN 978-0-521-54051-3 Paperback

9

Epipolar Geometry and the Fundamental Matrix

The epipolar geometry is the intrinsic projective geometry between two views. It is independent of scene structure, and only depends on the cameras' internal parameters and relative pose.

The fundamental matrix F encapsulates this intrinsic geometry. It is a 3 × 3 matrix of rank 2. If a point in 3-space **X** is imaged as **x** in the first view, and **x**′ in the second, then the image points satisfy the relation $\mathrm{x}'^{\mathsf{T}}\mathrm{F}\mathrm{x} = 0$.

We will first describe epipolar geometry, and derive the fundamental matrix. The properties of the fundamental matrix are then elucidated, both for general motion of the camera between the views, and for several commonly occurring special motions. It is next shown that the cameras can be retrieved from F up to a projective transformation of 3-space. This result is the basis for the projective reconstruction theorem given in chapter 10. Finally, if the camera internal calibration is known, it is shown that the Euclidean motion of the cameras between views may be computed from the fundamental matrix up to a finite number of ambiguities.

The fundamental matrix is independent of scene structure. However, it can be computed from correspondences of imaged scene points alone, without requiring knowledge of the cameras' internal parameters or relative pose. This computation is described in chapter 11.

9.1 Epipolar geometry

The epipolar geometry between two views is essentially the geometry of the intersection of the image planes with the pencil of planes having the baseline as axis (the baseline is the line joining the camera centres). This geometry is usually motivated by considering the search for corresponding points in stereo matching, and we will start from that objective here.

Suppose a point **X** in 3-space is imaged in two views, at **x** in the first, and **x**′ in the second. What is the relation between the corresponding image points **x** and **x**′? As shown in figure 9.1a the image points **x** and **x**′, space point **X**, and camera centres are coplanar. Denote this plane as ***π***. Clearly, the rays back-projected from **x** and **x** ′intersect at **X**, and the rays are coplanar, lying in ***π***. It is this latter property that is of most significance in searching for a correspondence.

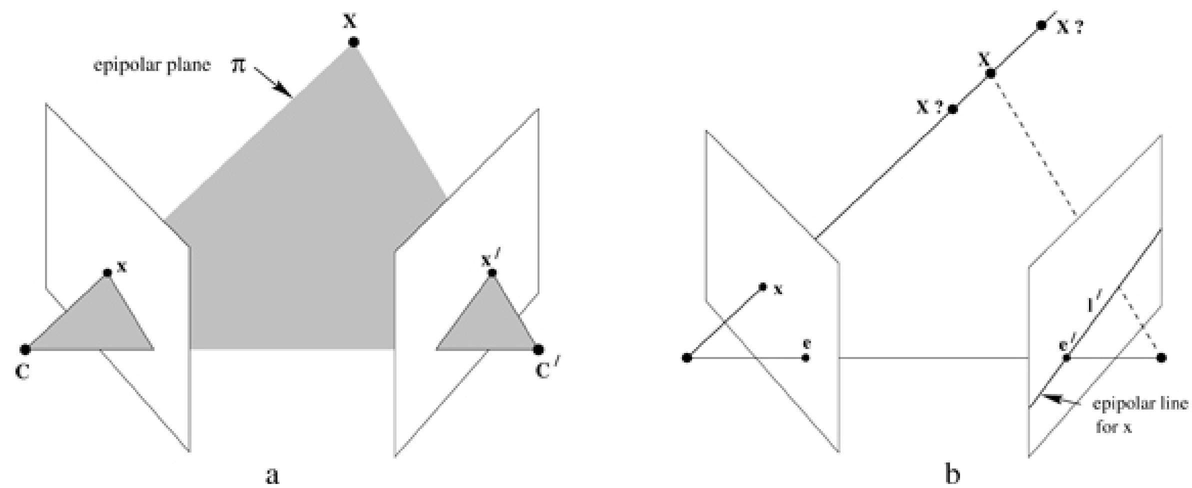


Fig. 9.1. **Point correspondence geometry.** *(a) The two cameras are indicated by their centres C and C′ and image planes. The camera centres, 3-space point X, and its images x and x′ lie in a common plane π. (b) An image point x back-projects to a ray in 3-space defined by the first camera centre, C, and x. This ray is imaged as a line l′ in the second view. The 3-space point X which projects to x must lie on this ray, so the image of X in the second view must lie on l′.*

Supposing now that we know only **x**, we may ask how the corresponding point **x**′ is constrained. The plane ***π*** is determined by the baseline and the ray defined by **x**. From above we know that the ray corresponding to the (unknown) point $\mathbf{x}'$ lies in $\boldsymbol{\pi}$, hence the point $\mathbf{x}'$ lies on the line of intersection $\mathbf{l}'$ of $\boldsymbol{\pi}$ with the second image plane. This line $\mathbf{l}'$ is the image in the second view of the ray back-projected from $\mathbf{x}$. It is the *epipolar line* corresponding to $\mathbf{x}$. In terms of a stereo correspondence algorithm the benefit is that the search for the point corresponding to $\mathbf{x}$ need not cover the entire image plane but can be restricted to the line $\mathbf{l}'$.

The geometric entities involved in epipolar geometry are illustrated in figure 9.2. The terminology is

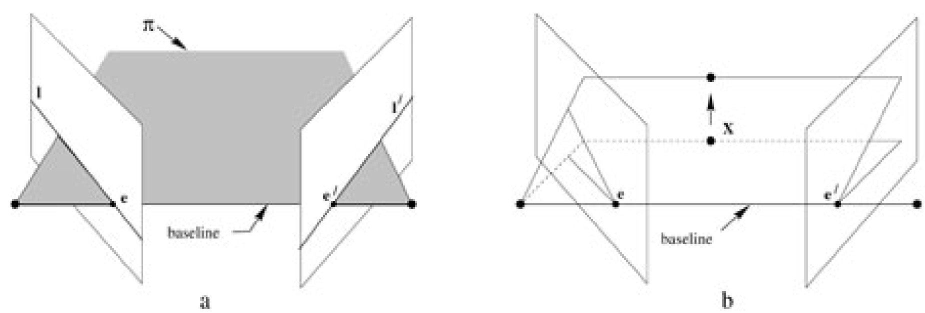


Fig. 9.2. **Epipolar geometry.** *(a) The camera baseline intersects each image plane at the epipoles e and e′. Any plane π containing the baseline is an epipolar plane, and intersects the image planes in corresponding epipolar lines l and l′. (b) As the position of the 3D point X varies, the epipolar planes "rotate" about the baseline. This family of planes is known as an epipolar pencil. All epipolar lines intersect at the epipole.*

- The **epipole** is the *point* of intersection of the line joining the camera centres (the baseline) with the image plane. Equivalently, the epipole is the image in one view of the camera centre of the other view. It is also the vanishing point of the baseline (translation) direction.
- An **epipolar plane** is a plane containing the baseline. There is a one-parameter family (a pencil) of epipolar planes.
- An **epipolar line** is the intersection of an epipolar plane with the image plane. All epipolar lines intersect at the epipole. An epipolar plane intersects the left and right image planes in

epipolar lines, and defines the correspondence between the lines.

Examples of epipolar geometry are given in figure 9.3 and figure 9.4. The epipolar geometry of these image pairs, and indeed all the examples of this chapter, is computed directly from the images as described in section 11.6(*p*290).

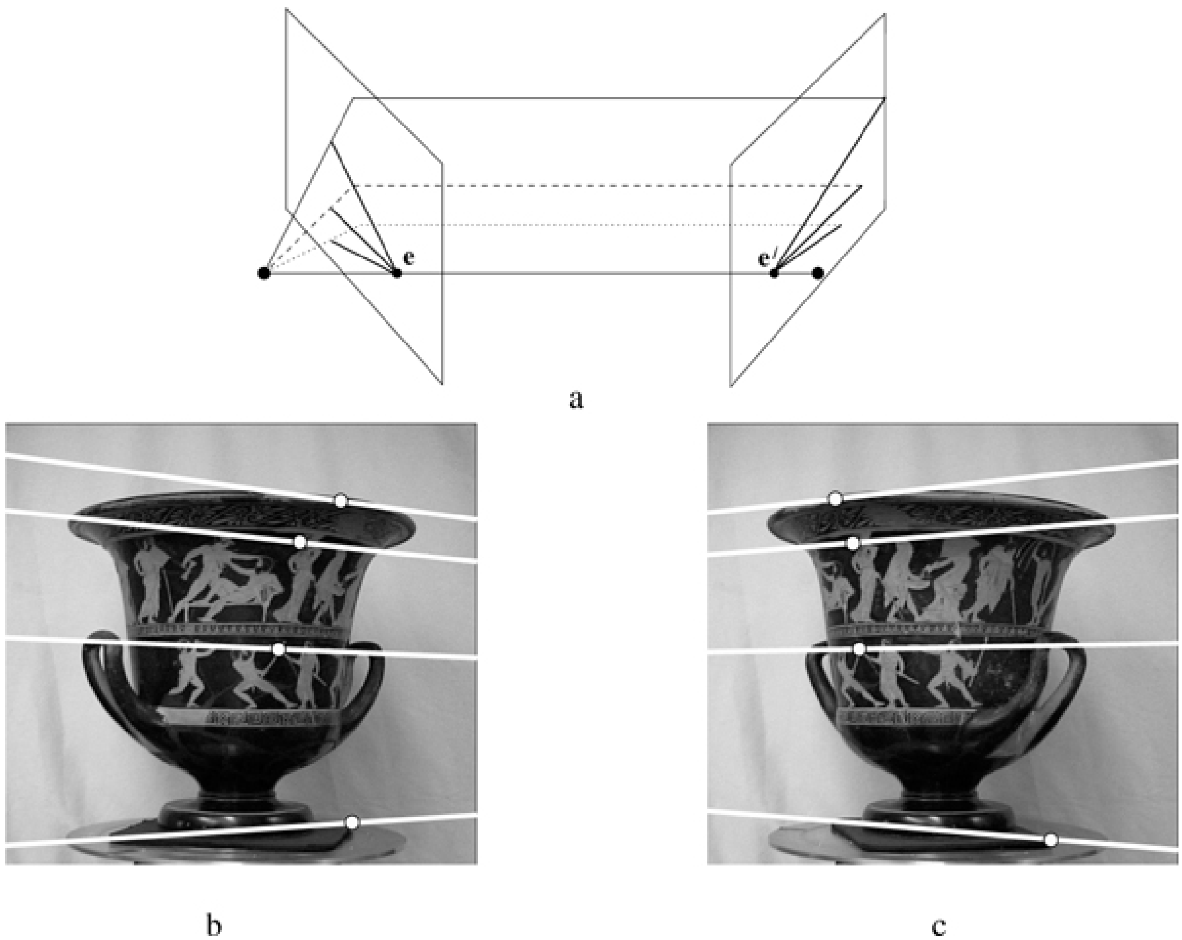


Fig. 9.3. **Converging cameras.** *(a) Epipolar geometry for converging cameras. (b) and (c) A pair of images with superimposed corresponding points and their epipolar lines (in white). The motion between the views is a translation and rotation. In each image, the direction of the other camera may be inferred from the intersection of the pencil of epipolar lines. In this case, both epipoles lie outside of the visible image.*

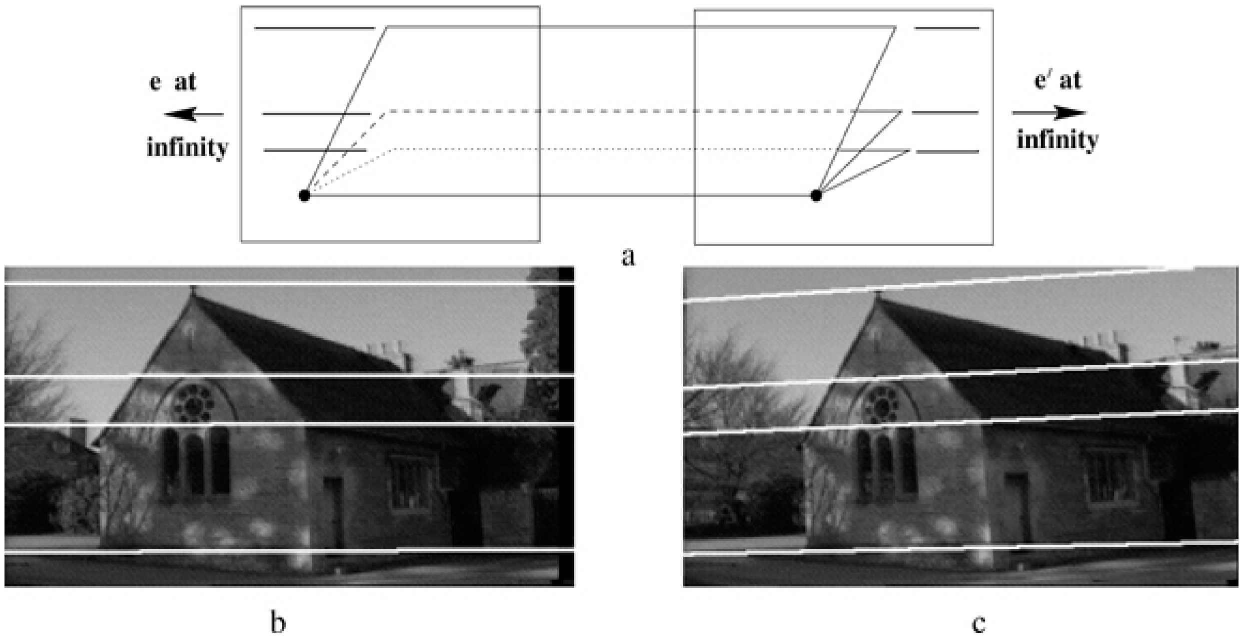


Fig. 9.4. **Motion parallel to the image plane.** *In the case of a special motion where the translation is parallel to the image plane, and the rotation axis is perpendicular to the image plane, the intersection of the baseline with the image plane is at infinity. Consequently the epipoles are at infinity, and epipolar lines are parallel. (a) Epipolar geometry for motion parallel to the image plane. (b) and (c) a pair of images for which the motion between views is (approximately) a translation parallel to the x-axis, with no rotation. Four corresponding epipolar lines are superimposed in white. Note that corresponding points lie on corresponding epipolar lines.*

9.2 The fundamental matrix F

The fundamental matrix is the algebraic representation of epipolar geometry. In the following we derive the fundamental matrix from the mapping between a point and its epipolar line, and then specify the properties of the matrix.

Given a pair of images, it was seen in figure 9.1 that to each point **x** in one image, there exists a corresponding epipolar line **l′** in the other image. Any point **x′** in the second image matching the point **x** must lie on the epipolar line **l′**. The epipolar line is the projection in the second image of the ray from the point **x** through the camera centre **C** of the first camera. Thus, there is a map

$$\mathbf{x} \mapsto \mathbf{l}'$$

from a point in one image to its corresponding epipolar line in the other image. It is the nature of this map that will now be explored. It will turn out that this mapping is a (singular) *correlation*, that is a projective mapping from points to lines, which is represented by a matrix F, the fundamental matrix.

9.2.1 Geometric derivation

We begin with a geometric derivation of the fundamental matrix. The mapping from a point in one image to a corresponding epipolar line in the other image may be decomposed into two steps. In the first step, the point **x** is mapped to some point **x**′ in the other image lying on the epipolar line **l**′. This point **x**′ is a potential match for the point **x**. In the second step, the epipolar line **l**′ is obtained as the line joining **x**′ to the epipole **e**′.

Step 1: Point transfer via a plane. Refer to figure 9.5. Consider a plane π in space not passing through either of the two camera centres. The ray through the first camera centre corresponding to the point **x** meets the plane π in a point **X**. This point **X** is then projected to a point **x**′ in the second image. This procedure is known as transfer via the plane π. Since **X** lies on the ray corresponding to **x**, the projected point **x**′ must lie on the epipolar line **l**′ corresponding to the image of this ray, as illustrated in figure 9.1b. The points **x** and **x**′ are both images of the 3D point **X** lying on a plane. The set of all such points $\mathbf{x}_i$ in the first image and the corresponding points $\mathbf{x}'_i$ in the second image are projectively equivalent, since they are each projectively equivalent to the planar point set $\mathbf{X}_i$. Thus there is a 2D homography H_π mapping each $\mathbf{x}_i$ to $\mathbf{x}'_i$.

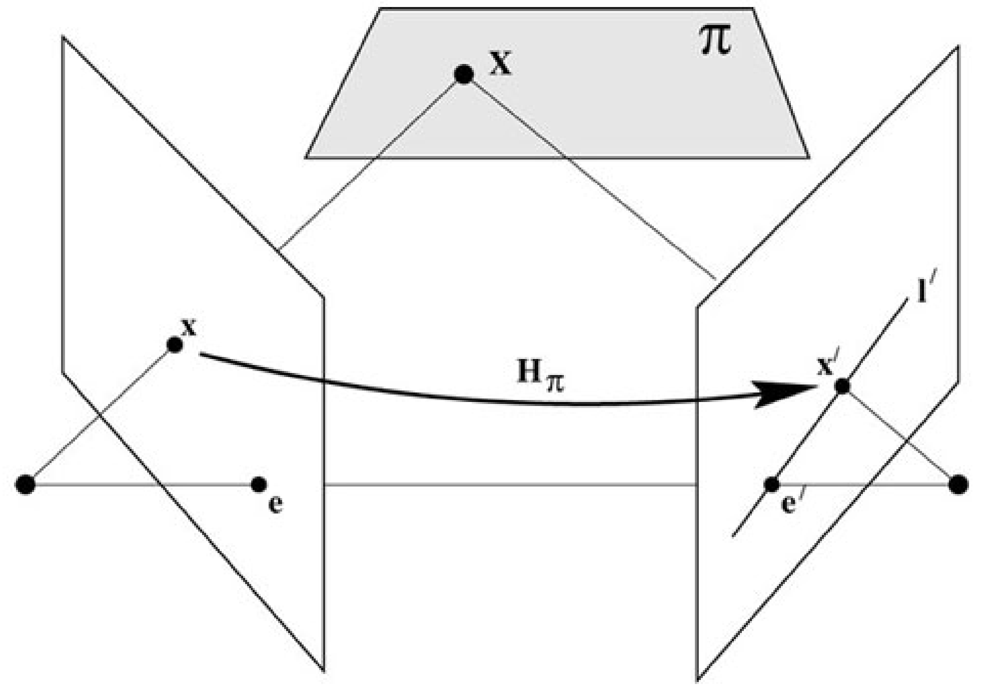


Fig. 9.5. *A point* **x** *in one image is transferred via the plane* ***π*** *to a matching point* **x** *in the second image. The epipolar line through* **x′** *is obtained by joining* **x′** *to the epipole* **e′**. *In symbols one may write* **x′** = H_π**x** *and* **l′** = [**e′**]$_\times$**x′** = [**e′**]$_\times H_\pi$**x** = F**x** *where* F = [**e′**]$_\times H_\pi$ *is the fundamental matrix*.

Step 2: Constructing the epipolar line. Given the point **x′** the epipolar line **l′** passing through **x′** and the epipole **e′** can be written as $l' = e' \times x' = [e']_\times x'$ (the notation $[e']_\times$ is defined in (A4.5–*p*581)). Since **x′** may be written as **x′** = H_πx, we have

$$l' = [e']_\times H_\pi x = Fx$$

where we define $F = [e']_\times H_\pi$, the fundamental matrix. This shows

Result 9.1. *The fundamental matrix* F *may be written as* $F = [e']_\times H_\pi$ *, where* H_π *is the transfer mapping from one image to another via any plane* ***π***. *Furthermore, since* $[e']_\times$ *has rank 2 and* H_π *rank 3*, F *is a matrix of rank 2.*

Geometrically, F represents a mapping from the 2-dimensional projective plane $\mathbb{P}^2$ of the first image to the pencil of epipolar lines through the epipole $\mathbf{e}'$. Thus, it represents a mapping from a 2-dimensional onto a 1-dimensional projective space, and hence must have rank 2.

Note, the geometric derivation above involves a scene plane $\boldsymbol{\pi}$, but a plane is *not* required in order for F to exist. The plane is simply used here as a means of defining a point map from one image to another. The connection between the fundamental matrix and transfer of points from one image to another via a plane is dealt with in some depth in chapter 13.

9.2.2 Algebraic derivation

The form of the fundamental matrix in terms of the two camera projection matrices, P, P′, may be derived algebraically. The following formulation is due to Xu and Zhang [Xu-96].

The ray back-projected from $\mathbf{x}$ by P is obtained by solving P$\mathbf{X}$ = $\mathbf{x}$. The one-parameter family of solutions is of the form given by (6.13–*p*162) as

$$\mathbf{X}(\lambda) = \mathrm{P}^{+}\mathbf{x} + \lambda\mathbf{C}$$

where P^{+} is the pseudo-inverse of P, i.e. $\mathrm{PP}^{+} = \mathrm{I}$, and $\mathbf{C}$ its null-vector, namely the camera centre, defined by P$\mathbf{C}$ = $\mathbf{0}$. The ray is parametrized by the scalar λ. In particular two points on the ray are $\mathrm{P}^{+}\mathbf{x}$ (at $\lambda = 0$), and the first camera centre $\mathbf{C}$ (at $\lambda = \infty$). These two points are imaged by the second camera P′ at $\mathrm{P}'\mathrm{P}^{+}\mathbf{x}$ and $\mathrm{P}'\mathbf{C}$ respectively in the second view. The epipolar line is the line joining these two projected points, namely $\mathbf{l}' = (\mathrm{P}'\mathbf{C}) \times (\mathrm{P}'\mathrm{P}^{+}\mathbf{x})$. The point $\mathrm{P}'\mathbf{C}$ is the epipole in the second image, namely the projection of the first camera centre, and may be denoted by $\mathbf{e}'$. Thus, $\mathbf{l}' = [\mathbf{e}']_{\times}(\mathrm{P}'\mathrm{P}^{+})\mathbf{x} = \mathrm{F}\mathbf{x}$, where F is the matrix

$$\mathrm{F} = [\mathbf{e}']_{\times}\mathrm{P}'\mathrm{P}^{+}. \tag{9.1}$$

This is essentially the same formula for the fundamental matrix as the one derived in the previous section, the homography H_π having the explicit form $H_\pi = P'P^+$ in terms of the two camera matrices. Note that this derivation breaks down in the case where the two camera centres are the same for, in this case, **C** is the common camera centre of both P and P′, and so P′**C** = **0**. It follows that F defined in (9.1) is the zero matrix.

Example 9.2. Suppose the camera matrices are those of a calibrated stereo rig with the world origin at the first camera

$$P = K[I \mid 0] \qquad P' = K'[R \mid t].$$

Then

$$P^+ = \begin{bmatrix} K^{-1} \\ 0^T \end{bmatrix} \qquad C = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

and

$$\begin{aligned} F &= [P'C]_\times P'P^+ \\ &= [K't]_\times K'RK^{-1} = K'^{-T}[t]_\times RK^{-1} = K'^{-T}R[R^Tt]_\times K^{-1} = K'^{-T}RK^T[KR^Tt]_\times \end{aligned}$$

(9.2

where the various forms follow from result A4.3(*p*582). Note that the epipoles (defined as the image of the other camera centre) are

$$e = P\begin{pmatrix} -R^Tt \\ 1 \end{pmatrix} = KR^Tt \qquad e' = P'\begin{pmatrix} 0 \\ 1 \end{pmatrix} = K't.$$

(9.3)

Thus we may write (9.2) as

$$F = [e']_\times K'RK^{-1} = K'^{-T}[t]_\times RK^{-1} = K'^{-T}R[R^T t]_\times K^{-1} = K'^{-T}RK^T[e]_\times.$$

(9.4

△

The expression for the fundamental matrix can be derived in many ways, and indeed will be derived again several times in this book. In particular, (17.3–*p*412) expresses F in terms of 4 × 4 determinants composed from rows of the camera matrices for each view.

9.2.3 Correspondence condition

Up to this point we have considered the map $\mathbf{x} \rightarrow \mathbf{l}'$ defined by F. We may now state the most basic properties of the fundamental matrix.

Result 9.3. *The fundamental matrix satisfies the condition that for any pair of corresponding points* $\mathbf{x} \leftrightarrow \mathbf{x}'$ *in the two images*

$$x'^T F x = 0.$$

This is true, because if points $\mathbf{x}$ and $\mathbf{x}'$ correspond, then $\mathbf{x}'$ lies on the epipolar line $\mathbf{l}' = F\mathbf{x}$ corresponding to the point $\mathbf{x}$. In other words $0 = x'^T l' = x'^T F x$. Conversely, if image points satisfy the relation $x'^T F x = 0$ then the rays defined by these points are coplanar. This is a necessary condition for points to correspond.

The importance of the relation of result 9.3 is that it gives a way of characterizing the fundamental matrix without reference to the camera matrices, i.e. only in terms of corresponding image points. This enables F to be computed from image correspondences alone. We have seen from (9.1) that F may be computed from the two camera matrices, P, P′, and in particular that F is determined uniquely from the cameras, up to an overall scaling. However, we may now enquire how many correspondences are required to compute F from $x'^T F x = 0$, and the circumstances under which the matrix is uniquely defined by these correspondences. The details of this are postponed until chapter 11, where it will be seen that in general at least 7 correspondences are required to compute F.

9.2.4 Properties of the fundamental matrix

Definition 9.4. Suppose we have two images acquired by cameras with non-coincident centres, then the **fundamental matrix** F is the unique 3×3 rank 2 homogeneous matrix which satisfies

$$x'^{\mathsf{T}} F x = 0 \tag{9.5}$$

for all corresponding points $\mathbf{x} \leftrightarrow \mathbf{x}'$.

We now briefly list a number of properties of the fundamental matrix. The most important properties are also summarized in table 9.1.

- F is a rank 2 homogeneous matrix with 7 degrees of freedom.
- **Point correspondence**: If **x** and **x** are corresponding image points, then
  $\mathbf{x}^{\mathsf{T}} F \mathbf{x} = 0$.
- **Epipolar lines**:
  - $\mathbf{l} = F\mathbf{x}$ is the epipolar line corresponding to **x**.
  - $\mathbf{l} = F^{\mathsf{T}}\mathbf{x}$ is the epipolar line corresponding to $\mathbf{x}'$.
- **Epipoles**:
  - $F\mathbf{e} = \mathbf{0}$.
  - $F^{\mathsf{T}}\mathbf{e}' = \mathbf{0}$.
- **Computation from camera matrices** P, P :
  - General cameras,
    $F = [\mathbf{e}\ ]_{\times} P\ P^{+}$, where P^{+} is the pseudo-inverse of P, and
    $\mathbf{e}' = P'\mathbf{C}$, with $P\mathbf{C} = \mathbf{0}$.
  - Canonical cameras, $P = [I\ /\ \mathbf{0}]$, $P' = [M\ /\ \mathbf{m}]$,
    $F = [\mathbf{e}']_{\times} M = M^{-\mathsf{T}}[\mathbf{e}]_{\times}$, where $\mathbf{e}' = \mathbf{m}$ and $\mathbf{e} = M^{-1}\mathbf{m}$.
  - Cameras not at infinity $P = K[I\ /\ \mathbf{0}]$, $P' = K'[R\ /\ \mathbf{t}]$,
    $F = K'^{-\mathsf{T}}[\mathbf{t}]_{\times} R K^{-1} = [K'\mathbf{t}]_{\times} K' R K^{-1} = K'^{-\mathsf{T}} R K^{\mathsf{T}} [K R^{\mathsf{T}} \mathbf{t}]_{\times}$.

Table 9.1. *Summary of fundamental matrix properties.*

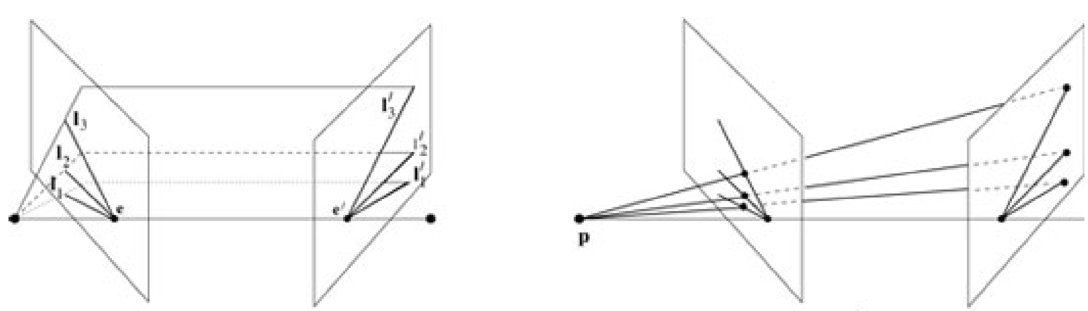
(i) **Transpose**: If F is the fundamental matrix of the pair of cameras (P, P′), then F^T is the fundamental matrix of the pair in the opposite order: (P′, P).

(ii) **Epipolar lines**: For any point **x** in the first image, the corresponding epipolar line is **l**′ = F**x**. Similarly, **l** = F^T**x**′ represents the epipolar line corresponding to **x**′ in the second image.

(iii) The **epipole**: for any point **x** (other than **e**) the epipolar line **l**′ = F**x** contains the epipole **e**′. Thus **e**′ satisfies $e'^{T}(Fx) = (e'^{T}F)x = 0$ for all **x**. It follows that $e'^{T}F = 0$, i.e. **e**′ is the left null-vector of F. Similarly F**e** = **0**, i.e. **e** is the right null-vector of F.

(iv) F has seven degrees of freedom: a 3×3 homogeneous matrix has eight independent ratios (there are nine elements, and the common scaling is not significant); however, F also satisfies the constraint det F = 0 which removes one degree of freedom.

(v) F is a *correlation*, a projective map taking a point to a line (see definition 2.29-(*p*59)). In this case a point in the first image **x** defines a line in the second **l**′ = F**x**, which is the epipolar line of **x**. If **l** and **l**′ are corresponding epipolar lines (see figure 9.6a) then any point **x** on **l** is mapped to the same line **l**′. This means there is no inverse mapping, and F is not of full rank. For this reason, F is not a proper correlation (which would be invertible).



a b

Fig. 9.6. **Epipolar line homography.** *(a) There is a pencil of epipolar lines in each image centred on the epipole. The correspondence between epipolar lines,* $\mathbf{l}_i \leftrightarrow \mathbf{l}'_i$*, is defined by the pencil of planes with axis the baseline. (b) The corresponding lines are related by a perspectivity with centre any point* **p** *on the baseline. It follows that the correspondence between epipolar lines in the pencils is a 1D homography.*

9.2.5 The epipolar line homography

The set of epipolar lines in each of the images forms a pencil of lines passing through the epipole. Such a pencil of lines may be considered as a 1-dimensional projective space. It is clear from figure 9.6b that corresponding epipolar lines are perspectively related, so that there is a homography between the pencil of epipolar lines centred at **e** in the first view, and the pencil centred at **e**′ in the second. A homography between two such 1-dimensional projective spaces has 3 degrees of freedom.

The degrees of freedom of the fundamental matrix can thus be counted as follows: 2 for **e**, 2 for **e**′, and 3 for the epipolar line homography which maps a line through **e** to a line through **e**′. A geometric representation of this homography is given in section 9.4. Here we give an explicit formula for this mapping.

Result 9.5. *Suppose* **l** *and* **l**′ *are corresponding epipolar lines, and* **k** *is any line not passing through the epipole* **e**, *then* **l** *and* **l**′ *are related by* $\mathbf{l}' = \mathrm{F}[\mathbf{k}]_\times \mathbf{l}$. *Symmetrically,* $\mathbf{l} = \mathrm{F}^\mathsf{T}[\mathbf{k}']_\times \mathbf{l}'$.

Proof. The expression $[\mathbf{k}]_\times \mathbf{l} = \mathbf{k} \times \mathbf{l}$ is the point of intersection of the two lines **k** and **l**, and hence a point on the epipolar line **l** – call it **x**. Hence, $\mathrm{F}[\mathbf{k}]_\times \mathbf{l} = \mathrm{F}\mathbf{x}$ is the epipolar line corresponding to the point **x**, namely the line **l**′.

□

Furthermore a convenient choice for **k** is the line **e**, since $\mathbf{k}^\mathsf{T}\mathbf{e} = \mathbf{e}^\mathsf{T}\mathbf{e} \neq 0$, so that the line **e** does not pass through the point **e** as is required. A similar argument holds for the choice of $\mathbf{k}' = \mathbf{e}'$. Thus the epipolar line homography may be written as

$$l' = F[e]_\times l \qquad l = F^T[e']_\times l' \ .$$

9.3 Fundamental matrices arising from special motions

A special motion arises from a particular relationship between the translation direction, **t**, and the direction of the rotation axis, **a**. We will discuss two cases: *pure translation*, where there is no rotation; and *pure planar motion*, where **t** is orthogonal to **a** (the significance of the planar motion case is described in section 3.4.1(*p*77)). The 'pure' indicates that there is no change in the internal parameters. Such cases are important, firstly because they occur in practice, for example a camera viewing an object rotating on a turntable is equivalent to planar motion for pairs of views; and secondly because the fundamental matrix has a special form and thus additional properties.

9.3.1 Pure translation

In considering pure translations of the camera, one may consider the equivalent situation in which the camera is stationary, and the world undergoes a translation −**t**. In this situation points in 3-space move on straight lines parallel to **t**, and the imaged intersection of these parallel lines is the vanishing point **v** in the direction of **t**. This is illustrated in figure 9.7 and figure 9.8. It is evident that **v** is the epipole for both views, and the imaged parallel lines are the epipolar lines. The algebraic details are given in the following example.

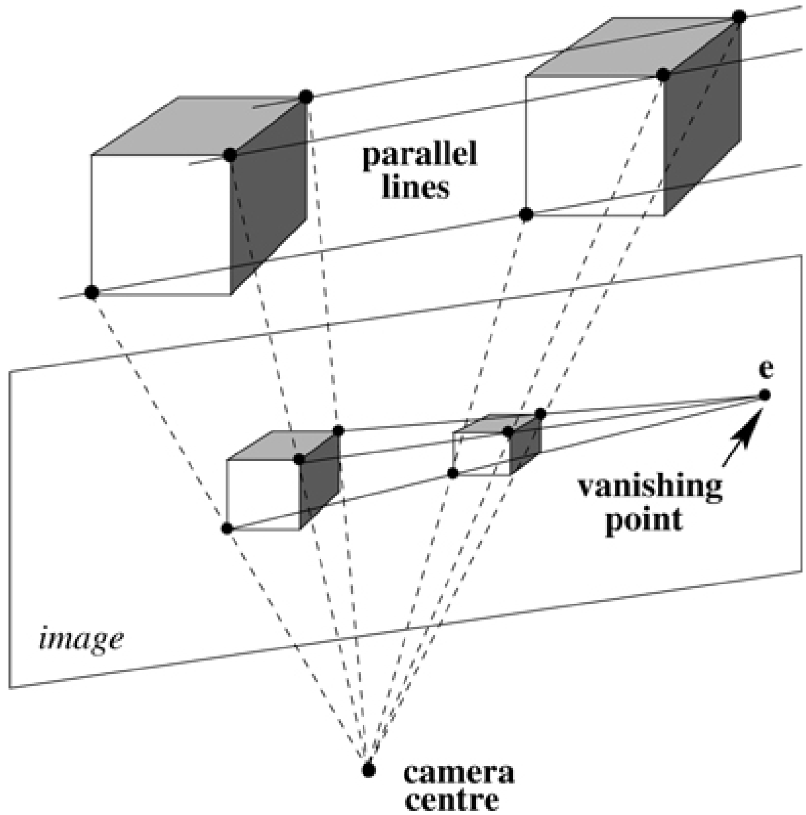


Fig. 9.7. *Under a pure translational camera motion, 3D points appear to slide along parallel rails. The images of these parallel lines intersect in a vanishing point corresponding to the translation direction. The epipole* **e** *is the vanishing point.*

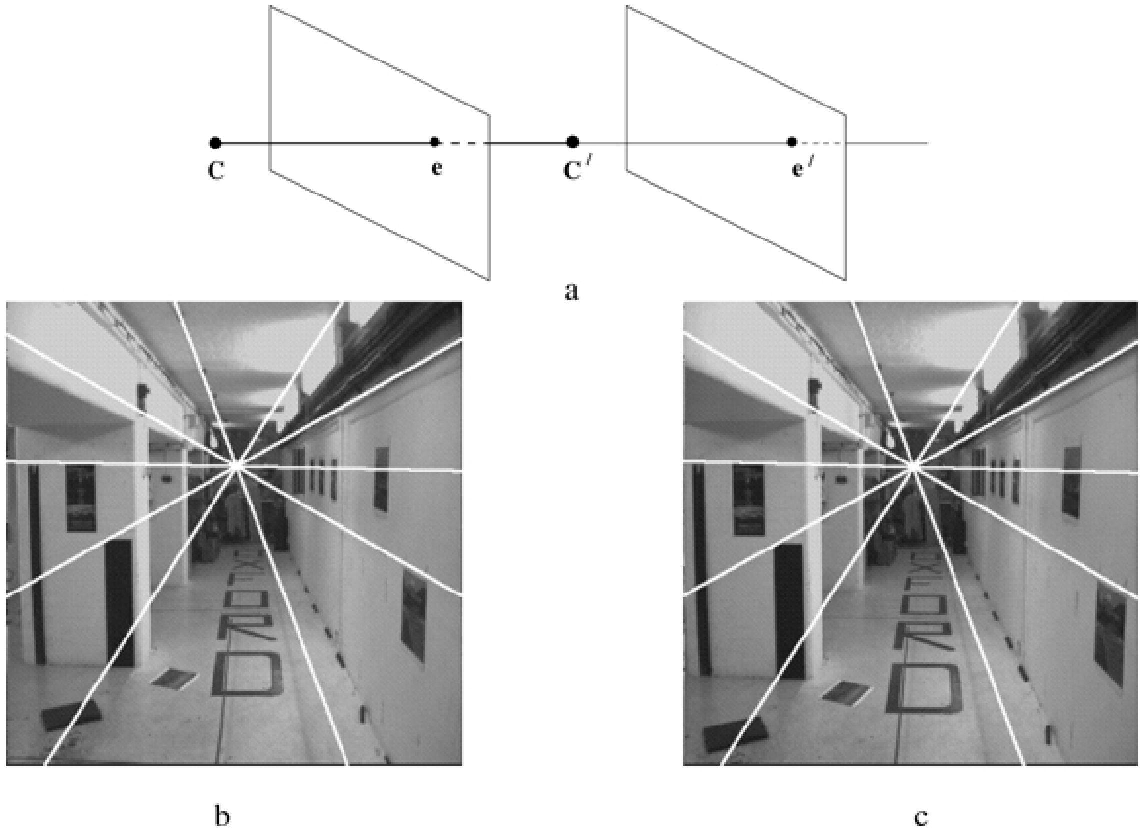


Fig. 9.8. **Pure translational motion.** *(a) under the motion the epipole is a fixed point, i.e. has the same coordinates in both images, and points appear to move along lines radiating from the epipole. The epipole in this case is termed the* Focus of Expansion *(FOE). (b) and (c) the same epipolar lines are overlaid in both cases. Note the motion of the posters on the wall which slide along the epipolar line.*

Example 9.6. Suppose the motion of the cameras is a pure translation with no rotation and no change in the internal parameters. One may assume that the two cameras are P = K[I / **0**] and P′ = K[I / **t**]. Then from (9.4) (using R = I and K = K′)

$$\mathrm{F} = [\mathrm{e}']_\times \mathrm{K}\mathrm{K}^{-1} = [\mathrm{e}']_\times .$$

If the camera translation is parallel to the *x*-axis, then $\mathbf{e}' = (1, 0, 0)^T$, so

$$F = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}.$$

The relation between corresponding points, $\mathbf{x}'^{\top} F \mathbf{x} = 0$, reduces to $y = y'$, i.e. the epipolar lines are corresponding rasters. This is the situation that is sought by image rectification described in section 11.12(*p*302).

△

Indeed if the image point $\mathbf{x}$ is normalized as $\mathbf{x} = (x, y, 1)^{\top}$, then from $\mathbf{x} = P\mathbf{X} = K[I \mid \mathbf{0}]\mathbf{X}$, the space point's (inhomogeneous) coordinates are $(X, Y, Z)^{\top} = ZK^{-1}\mathbf{x}$, where Z is the depth of the point $\mathbf{X}$ (the distance of $\mathbf{X}$ from the camera centre measured along the principal axis of the first camera). It then follows from $\mathbf{x}' = P'\mathbf{X} = K[I \mid \mathbf{t}]\mathbf{X}$ that the mapping from an image point $\mathbf{x}$ to an image point $\mathbf{x}'$ is

$$\mathbf{x}' = \mathbf{x} + K\mathbf{t}/Z. \tag{9.6}$$

The motion $\mathbf{x}' = \mathbf{x} + K\mathbf{t}/Z$ of (9.6) shows that the image point "starts" at $\mathbf{x}$ and then moves along the line defined by $\mathbf{x}$ and the epipole $\mathbf{e} = \mathbf{e}' = \mathbf{v}$. The extent of the motion depends on the magnitude of the translation $\mathbf{t}$ (which is not a homogeneous vector here) and the inverse depth Z, so that points closer to the camera appear to move faster than those further away – a common experience when looking out of a train window.

Note that in this case of pure translation $F = [\mathbf{e}']_{\times}$ is skew-symmetric and has only 2 degrees of freedom, which correspond to the position of the epipole. The epipolar line of $\mathbf{x}$ is $\mathbf{l}' = F\mathbf{x} = [\mathbf{e}]_{\times}\mathbf{x}$, and $\mathbf{x}$ lies on this line since $\mathbf{x}^{\top}[\mathbf{e}]_{\times}\mathbf{x} = 0$, i.e. $\mathbf{x}$, $\mathbf{x}'$ and $\mathbf{e} = \mathbf{e}'$ are collinear (assuming both images are overlaid on top of each other). This collinearity property is termed *auto-epipolar*, and does not hold for general motion.

General motion. The pure translation case gives additional insight into the general motion case. Given two arbitrary cameras, we may rotate the camera used for the first image so that it is aligned with the second camera. This rotation may be simulated by applying a projective transformation to the first image. A further correction may be applied to the first image to account for any difference in the calibration matrices of the two images. The result of these two corrections is a projective transformation H of the first image. If one assumes these corrections to have been made, then the effective relationship of the two cameras to each other is that of a pure translation. Consequently, the fundamental matrix corresponding to the corrected first image and the second image is of the form $\hat{F} = [e']_\times$, satisfying $x'^{\mathsf{T}}\hat{F}\hat{x} = 0$, where $\hat{x} = Hx$ is the corrected point in the first image. From this one deduces that $x'^{\mathsf{T}}[e']_\times Hx = 0$, and so the fundamental matrix corresponding to the initial point correspondences $\mathbf{x} \leftrightarrow \mathbf{x}'$ is $F = [\mathbf{e}']_\times H$. This is illustrated in figure 9.9.

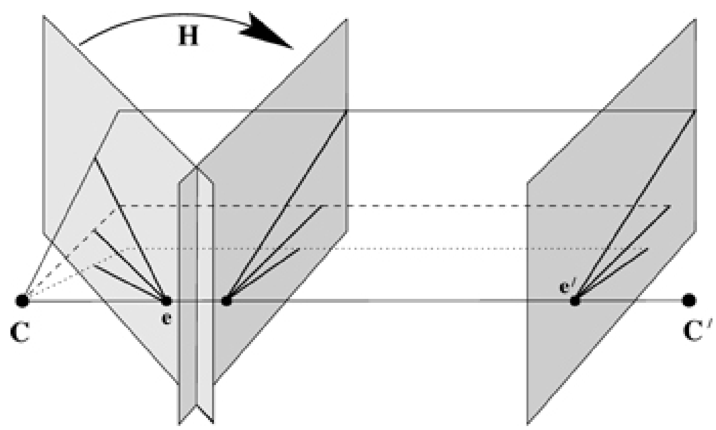


Fig. 9.9. **General camera motion.** *The first camera (on the left) may be rotated and corrected to simulate a pure translational motion. The fundamental matrix for the original pair is the product* $F = [\mathbf{e}']_\times H$, *where* $[\mathbf{e}']_\times$ *is the fundamental matrix of the translation, and* H *is the projective transformation corresponding to the correction of the first camera.*

Example 9.7. Continuing from example 9.2, assume again that the two cameras are $P = K[I \mid \mathbf{0}]$ and $P' = K'[R \mid \mathbf{t}]$. Then as described in section 8.4.2(*p*204) the requisite projective transformation is $H = K'RK^{-1} = H_\infty$, where H_∞ is the infinite homography (see section 13.4(*p*338)), and $F = [\mathbf{e}']_\times H_\infty$.

If the image point **x** is normalized as **x** = $(x, y, 1)^T$, as in example 9.6, then $(X, Y, Z)^T = ZK^{-1}\mathbf{x}$, and from **x** = P′**X** = K′[R / **t**]**X** the mapping from an image point **x** to an image point **x**′ is

$$x' = K'RK^{-1}x + K't/Z. \tag{9.7}$$

The mapping is in two parts: the first term depends on the image position alone, i.e. **x**, but not the point's depth Z, and takes account of the camera rotation and change of internal parameters; the second term depends on the depth, but not on the image position **x**, and takes account of camera translation. In the case of pure translation (R = I, K = K′) (9.7) reduces to (9.6).

△

9.3.2 Pure planar motion

In this case the rotation axis is orthogonal to the translation direction. Orthogonality imposes one constraint on the motion, and it is shown in the exercises at the end of this chapter that if K′ = K then F_s, the symmetric part of F, has rank 2 in this planar motion case (note, for a general motion the symmetric part of F has full rank). Thus, the condition that det $F_s = 0$ is an additional constraint on F and reduces the number of degrees of freedom from 7, for a general motion, to 6 degrees of freedom for a pure planar motion.

9.4 Geometric representation of the fundamental matrix

This section is not essential for a first reading and the reader may optionally skip to section 9.5.

In this section the fundamental matrix is decomposed into its symmetric and skew-symmetric parts, and each part is given a geometric representation. The symmetric and skew-symmetric parts of the fundamental matrix are

$$F_S = \left(F + F^T\right)/2 \qquad F_a = \left(F - F^T\right)/2$$

so that $F = F_s + F_a$.

To motivate the decomposition, consider the points $\mathbf{X}$ in 3-space that map to the same point in two images. These image points are fixed under the camera motion so that $\mathbf{x} = \mathbf{x}'$. Clearly such points are corresponding and thus satisfy $\mathbf{x}^T F \mathbf{x} = 0$, which is a necessary condition on corresponding points. Now, for any skew-symmetric matrix A the form $\mathbf{x}^T A \mathbf{x}$ is identically zero. Consequently only the symmetric part of F contributes to $\mathbf{x}^T F \mathbf{x} = 0$, which then reduces to $\mathbf{x}^T F_s \mathbf{x} = 0$. As will be seen below the matrix F_s may be thought of as a conic in the image plane.

Geometrically the conic arises as follows. The locus of all points in 3-space for which $\mathbf{x} = \mathbf{x}'$ is known as the *horopter* curve. Generally this is a twisted cubic curve in 3-space (see section 3.3(*p*75)) passing through the two camera centres [Maybank-93]. The image of the horopter is the conic defined by F_s. We return to the horopter in chapter 22.

Symmetric part. The matrix F_s is symmetric and is of rank 3 in general. It has 5 degrees of freedom and is identified with a point conic, called the *Steiner conic* (the name is explained below). The epipoles $\mathbf{e}$ and $\mathbf{e}'$ lie on the conic F_s. To see that the epipoles lie on the conic, i.e. that $\mathbf{e}^T F_s \mathbf{e} = 0$, start from $F\mathbf{e} = \mathbf{0}$. Then $\mathbf{e}^T F \mathbf{e} = \mathbf{0}$ and so $\mathbf{e}^T F_s \mathbf{e} + \mathbf{e}^T F_a \mathbf{e} = 0$. However, $\mathbf{e}^T F_a \mathbf{e} = 0$, since for any skew-symmetric matrix S, $\mathbf{x}^T S \mathbf{x} = 0$. Thus $\mathbf{e}^T F_s \mathbf{e} = 0$. The derivation for $\mathbf{e}'$ follows in a similar manner.

Skew-symmetric part. The matrix F_a is skew-symmetric and may be written as $F_a = [\mathbf{x}_a]_\times$, where $\mathbf{x}_a$ is the null-vector of F_a. The skew-symmetric part has 2 degrees of freedom and is identified with the point $\mathbf{x}_a$.

The relation between the point $\mathbf{x}_a$ and conic F_s is shown in figure 9.10a. The polar of $\mathbf{x}_a$ intersects the Steiner conic F_s at the epipoles $\mathbf{e}$ and $\mathbf{e}'$ (the pole–polar relation is described in section 2.2.3(*p*30)). The proof of this result is left as an exercise.

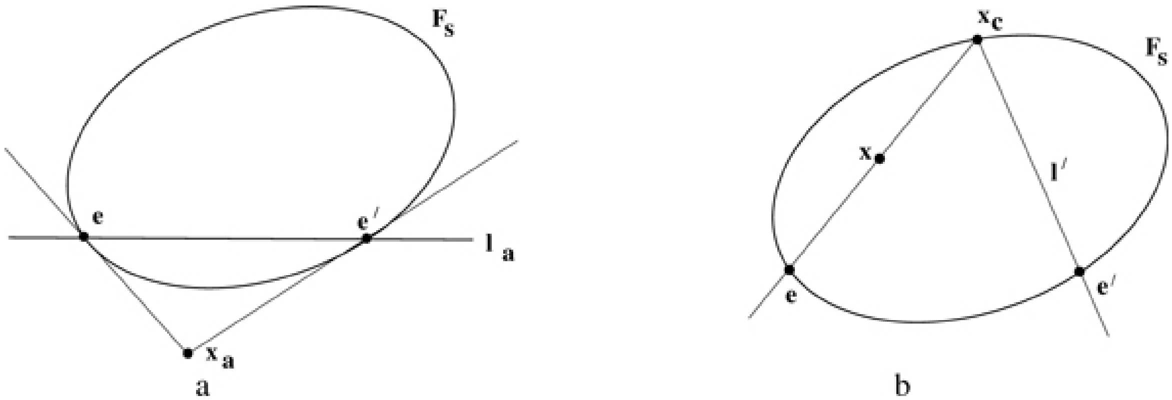


Fig. 9.10. **Geometric representation of** F. *(a) The conic* F_S *represents the symmetric part of* F, *and the point* $\mathbf{x}_a$ *the skew-symmetric part. The conic* F_S *is the locus of intersection of corresponding epipolar lines, assuming both images are overlaid on top of each other. It is the image of the horopter curve. The line* $\mathbf{l}_a$ *is the polar of* $\mathbf{x}_a$ *with respect to the conic* F_S*. It intersects the conic at the epipoles* **e** *and* **e**′. *(b) The epipolar line* **l**′ *corresponding to a point* **x** *is constructed as follows: intersect the line defined by the points* **e** *and* **x** *with the conic. This intersection point is* $\mathbf{x}_c$*. Then* **l**′ *is the line defined by the points* $\mathbf{x}_c$ *and* **e**′.

Epipolar line correspondence. It is a classical theorem of projective geometry due to Steiner [Semple-79] that for two line pencils related by a homography, the locus of intersections of corresponding lines is a conic. This is precisely the situation here. The pencils are the epipolar pencils, one through **e** and the other through **e**′. The epipolar lines are related by a 1D homography as described in section 9.2.5. The locus of intersection is the conic F_s.

The conic and epipoles enable epipolar lines to be determined by a geometric construction as illustrated in figure 9.10b. This construction is based on the fixed point property of the Steiner conic F_s. The epipolar line $\mathbf{l} = \mathbf{x} \times \mathbf{e}$ in the first view defines an epipolar plane in 3-space which intersects the horopter in a point, which we will call $\mathbf{X}_c$. The point $\mathbf{X}_c$ is imaged in the first view at $\mathbf{x}_c$, which is the point at which **l** intersects the conic F_s (since F_s is the image of the horopter). Now the image of $\mathbf{X}_c$ is also $\mathbf{x}_c$ in the second view due to the fixed-point property of the horopter. So $\mathbf{x}_c$ is the image in the second view of a point on the epipolar plane of **x**. It follows that $\mathbf{x}_c$ lies on the epipolar line $\mathbf{l}'$ of **x**, and consequently $\mathbf{l}'$ may be computed as $\mathbf{l}' = \mathbf{x}_c \times \mathbf{e}'$.

The conic together with two points on the conic account for the 7 degrees of freedom of F: 5 degrees of freedom for the conic and one each to specify the two epipoles on the conic. Given F, then the conic F_s, epipoles **e**, **e**′ and skew-symmetric point $\mathbf{x}_a$ are defined uniquely. However, F_s and $\mathbf{x}_a$ do not uniquely determine F since the identity of the epipoles is not recovered, i.e. the polar of $\mathbf{x}_a$ determines the epipoles but does not determine which one is **e** and which one **e**′.

9.4.1 Pure planar motion

We return to the case of planar motion discussed above in section 9.3.2, where F_s has rank 2. It is evident that in this case the Steiner conic is degenerate and from section 2.2.3(*p*30) is equivalent to two non-coincident lines:

$$F_S = l_h l_s^T + l_s l_h^T$$

as depicted in figure 9.11a. The geometric construction of the epipolar line $\mathbf{l}'$ corresponding to a point **x** of section 9.4 has a simple algebraic representation in this case. As in the general motion case, there are three steps, illustrated in figure 9.11b: first the line $\mathbf{l} = \mathbf{e} \times \mathbf{x}$ joining **e** and **x** is computed; second, its intersection point with the "conic" $\mathbf{x}_c = \mathbf{l}_s \times \mathbf{l}$ is determined; third the epipolar line $\mathbf{l}' = \mathbf{e}' \times \mathbf{x}_c$ is the join of $\mathbf{x}_c$ and **e**′. Putting these steps together we find

$$l' = e' \times [l_s \times (e \times x)] = [e']_\times [l_s]_\times [e]_\times x.$$

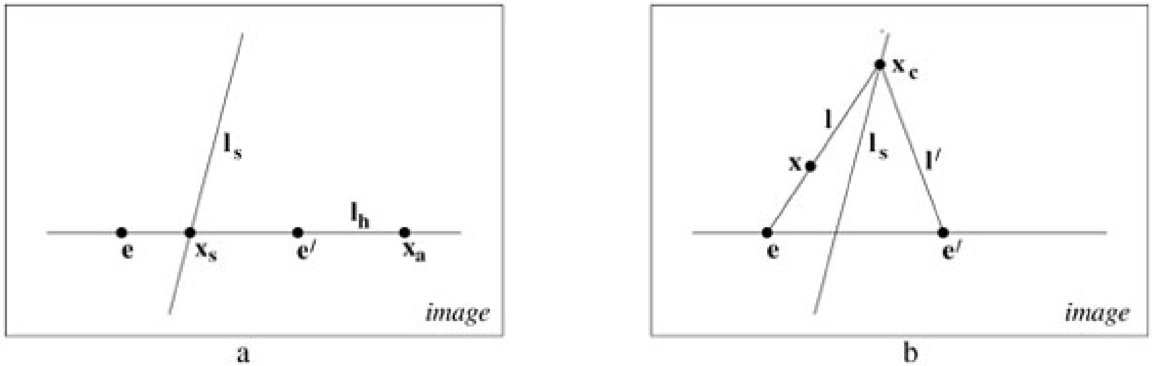


Fig. 9.11. **Geometric representation of** F **for planar motion.** *(a) The lines* $\mathbf{l}_s$ *and* $\mathbf{l}_h$ *constitute the Steiner conic for this motion, which is degenerate. Compare this figure with the conic for general motion shown in figure 9.10. (b) The epipolar line* $\mathbf{l}'$ *corresponding to a point* $\mathbf{x}$ *is constructed as follows: intersect the line defined by the points* $\mathbf{e}$ *and* $\mathbf{x}$ *with the (conic) line* $\mathbf{l}_s$*. This intersection point is* $\mathbf{x}_c$*. Then* $\mathbf{l}'$ *is the line defined by the points* $\mathbf{x}_c$ *and* $\mathbf{e}'$.

It follows that F may be written as

$$\mathrm{F} = [\mathrm{e}']_\times [\mathrm{l}_s]_\times [\mathrm{e}]_\times . \qquad (9.8)$$

The 6 degrees of freedom of F are accounted for as 2 degrees of freedom for each of the two epipoles and 2 degrees of freedom for the line.

The geometry of this situation can be easily visualized: the horopter for this motion is a degenerate twisted cubic consisting of a circle in the plane of the motion (the plane orthogonal to the rotation axis and containing the camera centres), and a line parallel to the rotation axis and intersecting the circle. The line is the screw axis (see section 3.4.1(*p*77)). The motion is equivalent to a rotation about the screw axis with zero translation. Under this motion points on the screw axis are fixed, and consequently their images are fixed. The line $\mathbf{l}_s$ is the image of the screw axis. The line $\mathbf{l}_h$ is the intersection of the image with the plane of the motion. This geometry is used for auto-calibration in chapter 19.

9.5 Retrieving the camera matrices

To this point we have examined the properties of F and of image relations for a point correspondence $\mathbf{x} \leftrightarrow \mathbf{x}'$. We now turn to one of the most significant properties of F, that the matrix may be used to determine the camera matrices of the two views.

9.5.1 Projective invariance and canonical cameras

It is evident from the derivations of section 9.2 that the map $\mathbf{l}' = F\mathbf{x}$ and the correspondence condition $\mathbf{x}'^{T}F\mathbf{x} = 0$ are *projective* relationships: the derivations have involved only projective geometric relationships, such as the intersection of lines and planes, and in the algebraic development only the linear mapping of the projective camera between world and image points. Consequently, the relationships depend only on projective co-ordinates in the image, and not, for example on Euclidean measurements such as the angle between rays. In other words the image relationships are projectively invariant: under a projective transformation of the image coordinates $\hat{x} = Hx, \hat{x}' = H'x'$, there is a corresponding map $\hat{l}' = \hat{F}\hat{x}$ with $\hat{F} = H'^{-T}FH^{-1}$ the corresponding rank 2 fundamental matrix.

Similarly, F only depends on projective properties of the cameras P, P′. The camera matrix relates 3-space measurements to image measurements and so depends on both the image coordinate frame and the choice of world coordinate frame. F does not depend on the choice of world frame, for example a rotation of world coordinates changes P, P′, but not F. In fact, the fundamental matrix is unchanged by a projective transformation of 3-space. More precisely,

Result 9.8. *If* H *is a* 4 × 4 *matrix representing a projective transformation of 3-space, then the fundamental matrices corresponding to the pairs of camera matrices* (P, P′) *and* (PH, P′H) *are the same.*

Proof. Observe that $P\mathbf{X} = (PH)(H^{-1}\mathbf{X})$, and similarly for P′. Thus if $\mathbf{x} \leftrightarrow \mathbf{x}'$ are matched points with respect to the pair of cameras (P, P′), corresponding to a 3D point $\mathbf{X}$, then they are also matched points with respect to the pair of cameras (PH, P′H), corresponding to the point $H^{-1}\mathbf{X}$.

□

Thus, although from (9.1–*p*244) a pair of camera matrices (P, P′) uniquely determine a fundamental matrix F, the converse is not true. The fundamental matrix determines the pair of camera matrices at best up to right-multiplication by a 3D projective transformation. It will be seen below that this is the full extent of the ambiguity, and indeed the camera matrices are determined up to a projective transformation by the fundamental matrix.

Canonical form of camera matrices. Given this ambiguity, it is common to define a specific *canonical form* for the pair of camera matrices corresponding to a given fundamental matrix in which the first matrix is of the simple form $[I \mid \mathbf{0}]$, where I is the 3×3 identity matrix and $\mathbf{0}$ a null 3-vector. To see that this is always possible, let P be augmented by one row to make a 4×4 non-singular matrix, denoted P^*. Now letting $H = P^{*-1}$, one verifies that $PH = [I \mid \mathbf{0}]$ as desired.

The following result is very frequently used

Result 9.9. *The fundamental matrix corresponding to a pair of camera matrices* $P = [I \mid \mathbf{0}]$ *and* $P' = [M \mid \mathbf{m}]$ *is equal to* $[\mathbf{m}]_\times M$.

This is easily derived as a special case of (9.1–*p*244).

9.5.2 Projective ambiguity of cameras given F

It has been seen that a pair of camera matrices determines a unique fundamental matrix. This mapping is not injective (one-to-one) however, since pairs of camera matrices that differ by a projective transformation give rise to the same fundamental matrix. It will now be shown that this is the only ambiguity. We will show that a given fundamental matrix determines the pair of camera matrices up to right multiplication by a projective transformation. Thus, the fundamental matrix captures the projective relationship of the two cameras.

Theorem 9.10. *Let* F *be a fundamental matrix and let* (P, P′) *and* ($\tilde{P}$, $\tilde{P}'$) *be two pairs of camera matrices such that* F *is the fundamental* *matrix corresponding to each of these pairs. Then there exists a non-singular* 4×4 *matrix* H *such that* $\tilde{P} = PH$ *and* $\tilde{P}' = P'H$.

Proof. Suppose that a given fundamental matrix F corresponds to two different pairs of camera matrices (P, P′) and ($\tilde{P}$, $\tilde{P}'$). As a first step, we may simplify the problem by assuming that each of the two pair of camera matrices is in canonical form with $P = \tilde{P} = [I \,/\, \mathbf{0}]$, since this may be done by applying projective transformations to each pair as necessary. Thus, suppose that $P = \tilde{P} = [I \,/\, \mathbf{0}]$ and that $P' = [A \,/\, \mathbf{a}]$ and $\tilde{P}' = [\tilde{A} \mid \tilde{a}]$. According to result 9.9 the fundamental matrix may then be written $F = [a]_\times A = [\tilde{a}]_\times \tilde{A}$.
We will need the following lemma:

Lemma 9.11. *Suppose the rank 2 matrix* F *can be decomposed in two different ways as* $F = [\mathbf{a}]_\times A$ *and* $F = [\tilde{a}]_\times \tilde{A}$*; then* $\tilde{a} = ka$ *and* $\tilde{A} = k^{-1}(A + av^\mathsf{T})$ *for some non-zero constant k and 3-vector* **v**.

Proof. First, note that $\mathbf{a}^\mathsf{T} F = \mathbf{a}^\mathsf{T}[\mathbf{a}]_\times A = \mathbf{0}$, and similarly, $\tilde{a}^\mathsf{T} F = 0$. Since F has rank 2, it follows that $\tilde{a} = ka$ as required. Next, from $[a]_\times A = [\tilde{a}]_\times \tilde{A}$ it follows that $[a]_\times \left(k\tilde{A} - A\right) = 0$, and so $k\tilde{A} - A = av^\mathsf{T}$ for some **v**. Hence, $\tilde{A} = k^{-1}(A + av^\mathsf{T})$ as required.

□

Applying this result to the two camera matrices P′ and $\tilde{P}'$ shows that $P' = [A \,/\, \mathbf{a}]$ and $\tilde{P}' = [k^{-1}(A + av^\mathsf{T}) \mid ka]$ if they are to generate the same F. It only remains now to show that these camera pairs are projectively related. Let H be the matrix $H = \begin{bmatrix} k^{-1}I & 0 \\ k^{-1}v^\mathsf{T} & k \end{bmatrix}$.

Then one verifies that $PH = k^{-1}[I \mid 0] = k^{-1}\tilde{P}$, and furthermore,

$$P'H = [A \mid a]H = [k^{-1}(A + av^\mathsf{T}) \mid ka] = [\tilde{A} \mid \tilde{a}] = \tilde{P}'$$

so that the pairs P, P′ and $\tilde{P}$, $\tilde{P}'$ are indeed projectively related.

□

This can be tied precisely to a counting argument: the two cameras P and P′ each have 11 degrees of freedom, making a total of 22 degrees of freedom. To specify a projective world frame requires 15 degrees of freedom (section 3.1(*p*65)), so once the degrees of freedom of the world frame are removed from the two cameras 22 − 15 = 7 degrees of freedom remain – which corresponds to the 7 degrees of freedom of the fundamental matrix.

9.5.3 Canonical cameras given F

We have shown that F determines the camera pair up to a projective transformation of 3-space. We will now derive a specific formula for a pair of cameras with canonical form given F. We will make use of the following characterization of the fundamental matrix F corresponding to a pair of camera matrices:

Result 9.12. *A non-zero matrix* F *is the fundamental matrix corresponding to a pair of camera matrices* P *and* P′ *if and only if* $P'^{T}FP$ *is skew-symmetric.*

Proof. The condition that $P'^{T}FP$ is skew-symmetric is equivalent to $\mathbf{X}^{T}P'^{T}FP\mathbf{X} = 0$ for all **X**. Setting $\mathbf{x}' = P'\mathbf{X}$ and $\mathbf{x} = P\mathbf{X}$, this is equivalent to $\mathbf{x}'^{T}F\mathbf{x} = 0$, which is the defining equation for the fundamental matrix.

□

One may write down a particular solution for the pairs of camera matrices in canonical form that correspond to a fundamental matrix as follows:

Result 9.13. *Let* F *be a fundamental matrix and* S *any skew-symmetric matrix. Define the pair of camera matrices*

$$P = [I \mid 0] \quad and \quad P' = [SF \mid e'],$$

where **e**′ *is the epipole such that* $\mathbf{e}'^{T}F = \mathbf{0}$, *and assume that* P′ *so defined is a valid camera matrix (has rank 3). Then* F *is the fundamental matrix corresponding to the pair* (P, P′).

To demonstrate this, we invoke result 9.12 and simply verify that

$$[\mathrm{SF} \mid \mathrm{e}']^{\mathsf{T}}\mathrm{F}[\mathrm{I} \mid 0] = \begin{bmatrix} \mathrm{F}^{\mathsf{T}}\mathrm{S}^{\mathsf{T}}\mathrm{F} & 0 \\ \mathrm{e}'^{\mathsf{T}}\mathrm{F} & 0 \end{bmatrix} = \begin{bmatrix} \mathrm{F}^{\mathsf{T}}\mathrm{S}^{\mathsf{T}}\mathrm{F} & 0 \\ 0^{\mathsf{T}} & 0 \end{bmatrix} \tag{9.9}$$

which is skew-symmetric.

The skew-symmetric matrix S may be written in terms of its null-vector as $\mathrm{S} = [\mathbf{s}]_\times$. Then $[[\mathbf{s}]_\times \mathrm{F} \mid \mathbf{e}']$ has rank 3 provided $\mathbf{s}^{\mathsf{T}}\mathbf{e}' \neq 0$, according to the following argument. Since $\mathbf{e}'\mathrm{F} = \mathbf{0}$, the column space (span of the columns) of F is perpendicular to $\mathbf{e}'$. But if $\mathbf{s}^{\mathsf{T}}\mathbf{e}' = 0$, then $\mathbf{s}$ is not perpendicular to $\mathbf{e}'$, and hence not in the column space of F. Now, the column space of $[\mathbf{s}]_\times\mathrm{F}$ is spanned by the cross-products of $\mathbf{s}$ with the columns of F, and therefore equals the plane perpendicular to $\mathbf{s}$. So $[\mathbf{s}]_\times\mathrm{F}$ has rank 2. Since $\mathbf{e}'$ is not perpendicular to $\mathbf{s}$, it does not lie in this plane, and so $[[\mathbf{s}]_\times\mathrm{F} \mid \mathbf{e}']$ has rank 3, as required.

As suggested by Luong and Viéville [Luong-96] a good choice for S is $\mathrm{S} = [\mathbf{e}']_\times$, for in this case $\mathbf{e}'^{\mathsf{T}}\mathbf{e}' \neq 0$, which leads to the following useful result.

Result 9.14. *The camera matrices corresponding to a fundamental matrix* F *may be chosen as* $\mathrm{P} = [\mathrm{I} \mid \mathbf{0}]$ *and* $\mathrm{P}' = [[\mathbf{e}']_\times\mathrm{F} \mid \mathbf{e}']$.

Note that the camera matrix P′ has left 3 × 3 submatrix $[\mathbf{e}']_\times\mathrm{F}$ which has rank 2. This corresponds to a camera with centre on $\boldsymbol{\pi}_\infty$. However, there is no particular reason to avoid this situation.

The proof of theorem 9.10 shows that the four parameter family of camera pairs in canonical form $\tilde{\mathrm{P}} = [\mathrm{I} \mid 0]$, $\tilde{\mathrm{P}}' = [\mathrm{A} + \mathrm{a}\mathrm{v}^{\mathsf{T}} \mid k\mathrm{a}]$ have the same fundamental matrix as the canonical pair, $\mathrm{P} = [\mathrm{I} \mid \mathbf{0}]$, $\mathrm{P}' = [\mathrm{A} \mid \mathbf{a}]$; and that this is the most general solution. To summarize:

Result 9.15. *The general formula for a pair of canonic camera matrices corresponding to a fundamental matrix* F *is given by*

$$P = [I \mid 0] \quad P' = [[e']_\times F + e'v^T \mid \lambda e']$$

(9.10)

where **v** *is any 3-vector, and* λ *a non-zero scalar.*

9.6 The essential matrix

The essential matrix is the specialization of the fundamental matrix to the case of normalized image coordinates (see below). Historically, the essential matrix was introduced (by Longuet-Higgins) before the fundamental matrix, and the fundamental matrix may be thought of as the generalization of the essential matrix in which the (inessential) assumption of calibrated cameras is removed. The essential matrix has fewer degrees of freedom, and additional properties, compared to the fundamental matrix. These properties are described below.

Normalized coordinates. Consider a camera matrix decomposed as P = K[R / **t**], and let **x** = P**X** be a point in the image. If the calibration matrix K is known, then we may apply its inverse to the point **x** to obtain the point $\hat{x}$ = K^{-1}**x**. Then $\hat{x}$ = [R / **t**]**X**, where $\hat{x}$ is the image point expressed in *normalized coordinates*. It may be thought of as the image of the point **X** with respect to a camera [R / **t**] having the identity matrix I as calibration matrix. The camera matrix K^{-1}P = [R / **t**] is called a *normalized camera matrix*, the effect of the known calibration matrix having been removed.

Now, consider a pair of normalized camera matrices P = [I / **0**] and P′ = [R / **t**]. The fundamental matrix corresponding to the pair of normalized cameras is customarily called the *essential matrix*, and according to (9.2–*p*244) it has the form

$$E = [t]_\times R = R\,[R^T t]_\times .$$

Definition 9.16. The defining equation for the essential matrix is

$$\hat{x}'^T E \hat{x} = 0 \tag{9.11}$$

in terms of the normalized image coordinates for corresponding points $\mathbf{x} \leftrightarrow \mathbf{x}'$.

Substituting for $\hat{x}$ and $\hat{x}'$ gives $\mathbf{x}'^{T}K'^{-T}EK^{-1}\mathbf{x} = 0$. Comparing this with the relation $\mathbf{x}'^{T}F\mathbf{x} = 0$ for the fundamental matrix, it follows that the relationship between the fundamental and essential matrices is

$$E = K'^{T}FK. \tag{9.12}$$

9.6.1 Properties of the essential matrix

The essential matrix, $E = [\mathbf{t}]_{\times}R$, has only five degrees of freedom: both the rotation matrix R and the translation $\mathbf{t}$ have three degrees of freedom, but there is an overall scale ambiguity – like the fundamental matrix, the essential matrix is a homogeneous quantity.

The reduced number of degrees of freedom translates into extra constraints that are satisfied by an essential matrix, compared with a fundamental matrix. We investigate what these constraints are.

Result 9.17. *A* 3×3 *matrix is an essential matrix if and only if two of its singular values are equal, and the third is zero.*

Proof. This is easily deduced from the decomposition of E as $[\mathbf{t}]_{\times}R = SR$, where S is skew-symmetric. We will use the matrices

$$W = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad Z = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{9.13}$$

It may be verified that W is orthogonal and Z is skew-symmetric. From Result A4.1-(*p*581), which gives a block decomposition of a general skew-symmetric matrix, the 3 × 3 skew-symmetric matrix S may be written as $S = kUZU^{T}$ where U is orthogonal.

Noting that, up to sign, $Z = \text{diag}(1, 1, 0)W$, then up to scale, $S = U\,\text{diag}(1, 1, 0)WU^{T}$, and $E = SR = U\,\text{diag}(1, 1, 0)(WU^{T}R)$. This is a singular value decomposition of E with two equal singular values, as required. Conversely, a matrix with two equal singular values may be factored as SR in this way.

□

Since $E = U \operatorname{diag}(1, 1, 0)V^T$, it may seem that E has six degrees of freedom and not five, since both U and V have three degrees of freedom. However, because the two singular values are equal, the SVD is not unique – in fact there is a one-parameter family of SVDs for E. Indeed, an alternative SVD is given by $E = (U \operatorname{diag}(R_{2\times2}, 1)) \operatorname{diag}(1, 1, 0)(\operatorname{diag}(R^T_{2\times2}, 1))V^T$ for any 2 × 2 rotation matrix R.

9.6.2 Extraction of cameras from the essential matrix

The essential matrix may be computed directly from (9.11) using normalized image coordinates, or else computed from the fundamental matrix using (9.12). (Methods of computing the fundamental matrix are deferred to chapter 11). Once the essential matrix is known, the camera matrices may be retrieved from E as will be described next. In contrast with the fundamental matrix case, where there is a projective ambiguity, the camera matrices may be retrieved from the essential matrix up to scale and a four-fold ambiguity. That is there are four possible solutions, except for overall scale, which cannot be determined.

We may assume that the first camera matrix is $P = [I \mid \mathbf{0}]$. In order to compute the second camera matrix, P′, it is necessary to factor E into the product SR of a skew-symmetric matrix and a rotation matrix.

Result 9.18. *Suppose that the SVD of* E *is* $U \operatorname{diag}(1, 1, 0)V^T$*. Using the notation of (9.13), there are (ignoring signs) two possible factorizations* E = SR *as follows:*

$$S = UZU^T \quad R = UWV^T \;\; or \;\; UW^TV^T \;. \tag{9.14}$$

Proof. That the given factorization is valid is true by inspection. That there are no other factorizations is shown as follows. Suppose E = SR. The form of S is determined by the fact that its left null-space is the same as that of E. Hence $S = UZU^T$. The rotation R may be written as UXV^T, where X is some rotation matrix. Then

$$U\,\mathrm{diag}(1,1,0)V^T = E = SR = (UZU^T)(UXV^T) = U(ZX)V^T$$

from which one deduces that ZX = diag(1, 1, 0). Since X is a rotation matrix, it follows that X = W or $X = W^T$, as required.

□

The factorization (9.14) determines the **t** part of the camera matrix P′, up to scale, from $S = [\mathbf{t}]_\times$. However, the Frobenius norm of $S = UZU^T$ is $\sqrt{2}$, which means that if $S = [\mathbf{t}]_\times$ *including scale* then $||\mathbf{t}|| = 1$, which is a convenient normalization for the baseline of the two camera matrices. Since $S\mathbf{t} = \mathbf{0}$, it follows that $\mathbf{t} = U\,(0, 0, 1)^T = \mathbf{u}_3$, the last column of U. However, the sign of E, and consequently **t**, cannot be determined. Thus, corresponding to a given essential matrix, there are four possible choices of the camera matrix P′, based on the two possible choices of R and two possible signs of t. To summarize:

Result 9.19. *For a given essential matrix* $E = U\,\mathrm{diag}(1, 1, 0)V^T$, *and first camera matrix* $P = [I \mid \mathbf{0}]$, *there are four possible choices for the second camera matrix* P′, *namely*

$$P' = [UWV^T \mid +\mathbf{u}_3] \text{ or } [UWV^T \mid -\mathbf{u}_3] \text{ or } [UW^TV^T \mid +\mathbf{u}_3] \text{ or } [UW^TV^T \mid -\mathbf{u}_3]$$

9.6.3 Geometrical interpretation of the four solutions

It is clear that the difference between the first two solutions is simply that the direction of the translation vector from the first to the second camera is reversed.

The relationship of the first and third solutions in result 9.19 is a little more complicated. However, it may be verified that

$$[UW^TV^T \mid \mathbf{u}_3] = [UWV^T \mid \mathbf{u}_3]\begin{bmatrix} VW^TW^TV^T & \\ & 1 \end{bmatrix}$$

and $VW^TW^TV^T = V\,\mathrm{diag}(-1,-1,1)V^T$ is a rotation through 180° about the line joining the two camera centres. Two solutions related in this way are known as a "twisted pair".

The four solutions are illustrated in figure 9.12, where it is shown that a reconstructed point **X** will be in front of both cameras in one of these four solutions only. Thus, testing with a single point to determine if it is in front of both cameras is sufficient to decide between the four different solutions for the camera matrix P′.

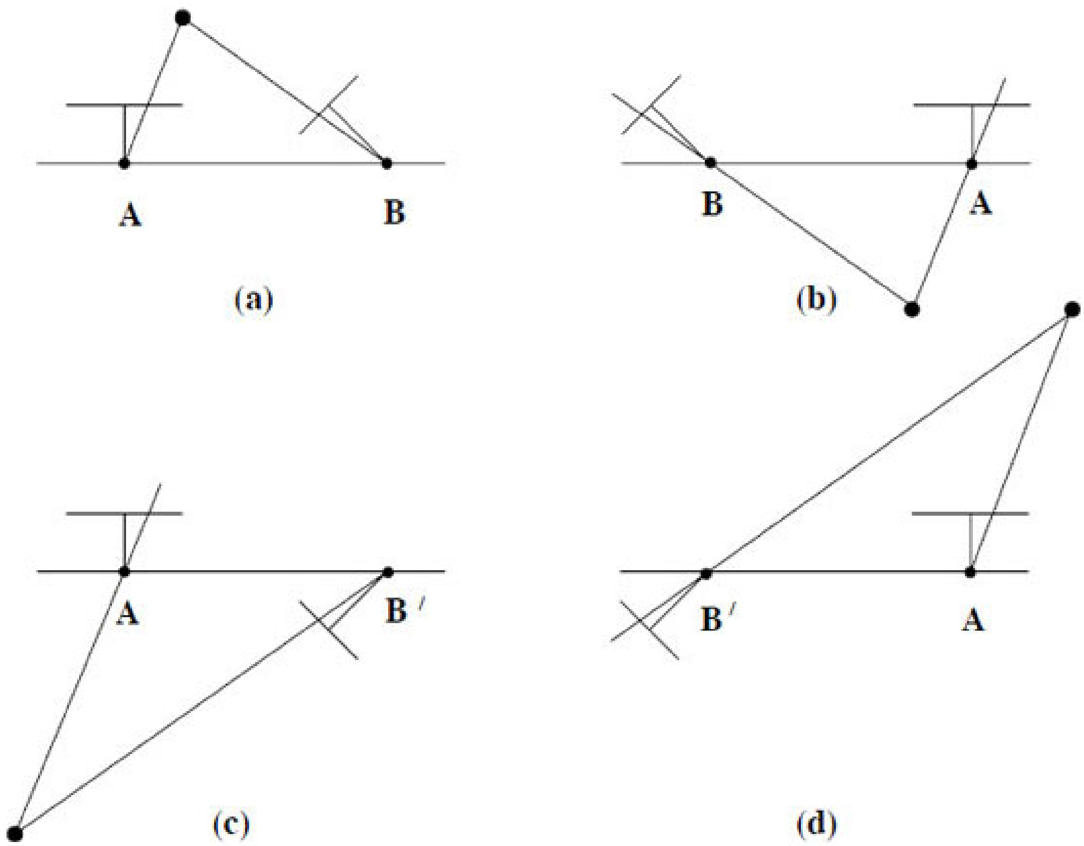


Fig. 9.12. **The four possible solutions for calibrated reconstruction from** E. *Between the left and right sides there is a baseline reversal. Between the top and bottom rows camera B rotates* 180° *about the baseline. Note, only in (a) is the reconstructed point in front of both cameras.*

Note. The point of view has been taken here that the essential matrix is a homogeneous quantity. An alternative point of view is that the essential matrix is defined exactly by the equation $E = [\mathbf{t}]_\times R$, (i.e.

including scale), and is determined only up to indeterminate scale by the equation $\mathbf{x}'^{\mathsf{T}}\mathrm{E}\mathbf{x} = 0$. The choice of point of view depends on which of these two equations one regards as the defining property of the essential matrix.

9.7 Closure

9.7.1 The literature

The essential matrix was introduced to the computer vision community by Longuet-Higgins [LonguetHiggins-81], with a matrix analogous to E appearing in the photogrammetry literature, e.g. [VonSanden-08]. Many properties of the essential matrix have been elucidated particularly by Huang and Faugeras [Huang-89], [Maybank-93], and [Horn-90].

The realization that the essential matrix could also be applied in uncalibrated situations, as it represented a projective relation, developed in the early part of the 1990s, and was published simultaneously by Faugeras [Faugeras-92b, Faugeras-92a], and Hartley *et al.* [Hartley-92a, Hartley-92c].

The special case of pure planar motion was examined by [Maybank-93] for the essential matrix. The corresponding case for the fundamental matrix is investigated by Beardsley and Zisserman [Beardsley-95a] and Viéville and Lingrand [Vieville-95], where further properties are given.

9.7.2 Notes and exercises

(i) **Fixating cameras.** Suppose two cameras fixate on a point in space such that their principal axes intersect at that point. Show that if the image coordinates are normalized so that the coordinate origin coincides with the principal point then the F_{33} element of the fundamental matrix is zero.

(ii) **Mirror images.** Suppose that a camera views an object and its reflection in a plane mirror. Show that this situation is equivalent to two views of the object, and that the fundamental matrix is skew-symmetric. Compare the fundamental matrix for this configuration with that of: (a) a pure translation, and (b) a pure planar motion. Show that the fundamental matrix is auto-epipolar (as is (a)).

(iii) Show that if the vanishing line of a plane contains the epipole then the plane is parallel to the baseline.

(iv) **Steiner conic.** Show that the polar of $\mathbf{x}_a$ intersects the Steiner conic F_s at the epipoles (figure 9.10a). Hint, start from $F\mathbf{e} = F_s\mathbf{e} + F_a\mathbf{e} = \mathbf{0}$. Since $\mathbf{e}$ lies on the conic F_s, then $\mathbf{l}_1 = F_s\mathbf{e}$ is the tangent line at $\mathbf{e}$, and $\mathbf{l}_2 = F_a\mathbf{e} = [\mathbf{x}_a]_\times\mathbf{e} = \mathbf{x}_a \times \mathbf{e}$ is a line through $\mathbf{x}_a$ and $\mathbf{e}$.

(v) The affine type of the Steiner conic (hyperbola, ellipse or parabola as given in section 2.8.2(*p*59)) depends on the relative configuration of the two cameras. For example, if the two cameras are facing each other then the Steiner conic is a hyperbola. This is shown in [Chum-03] where further results on oriented epipolar geometry are given.

(vi) **Planar motion.** It is shown by [Maybank-93] that if the rotation axis direction is orthogonal or parallel to the translation direction then the symmetric part of the essential matrix has rank 2. We assume here that $K = K'$. Then from (9.12), $F = K^{-T}EK^{-1}$, and so

$$F_S = (F + F^T)/2 = K^{-T}(E + E^T)K^{-1}/2 = K^{-T}E_SK^{-1}.$$

It follows from $\det(F_s) = \det(K^{-1})^2 \det(E_s)$ that the symmetric part of F is also singular. Does this result hold if $K \neq K'$?

(vii) Any matrix F of rank 2 is the fundamental matrix corresponding to some pair of camera matrices (P, P′) This follows directly from result 9.14 since the solution for the canonical cameras depends only on the rank 2 property of F.

(viii) Show that the 3D points determined from one of the ambiguous reconstructions obtained from E are related to the corresponding 3D points determined from another reconstruction by either (i) an inversion through the second camera centre; or (ii) a harmonic homology of 3-space (see section A7.2(*p*629)), where the homology plane is perpendicular to the baseline and through the second camera centre, and the vertex is the first camera centre.

(ix) Following a similar development to section 9.2.2, derive the form of the fundamental matrix for two linear pushbroom cameras. Details of this matrix are given in [Gupta-97] where it is shown that affine reconstruction is possible from a pair of images.

13

Scene planes and homographies

This chapter describes the projective geometry of two cameras and a world plane.

Images of points on a plane are related to corresponding image points in a second view by a (planar) homography as shown in figure 13.1. This is a projective relation since it depends only on the intersections of planes with lines. It is said that the plane *induces* a homography between the views. The homography map *transfers* points from one view to the other as if they were images of points on the plane.

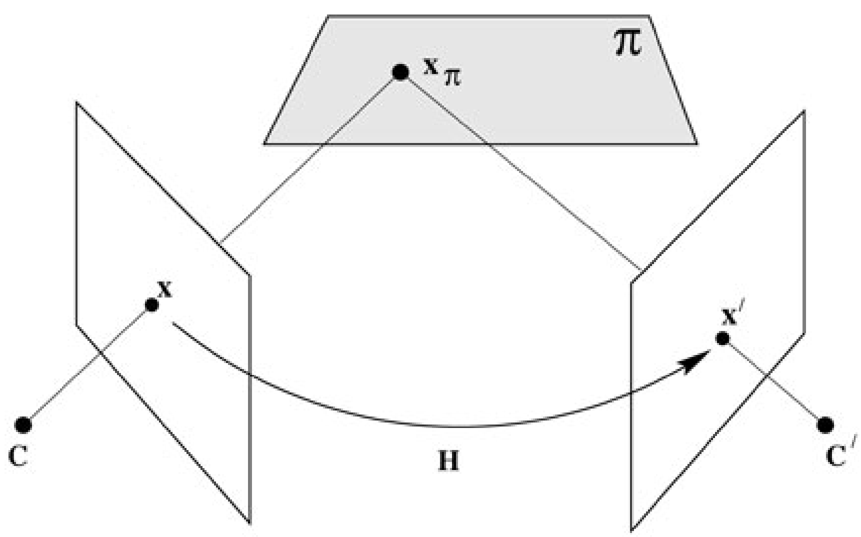


Fig. 13.1. **The homography induced by a plane.** *The ray corresponding to a point* $\mathbf{x}$ *is extended to meet the plane* $\boldsymbol{\pi}$ *in a point* $\mathbf{x}_{\pi}$*; this point is projected to a point* $\mathbf{x}'$ *in the other image. The map from* $\mathbf{x}$ *to* $\mathbf{x}'$ *is the homography induced*

by the plane $\boldsymbol{\pi}$. There is a perspectivity, $\mathbf{x} = \mathrm{H}_{1\pi}\mathbf{x}_{\pi}$, *between the world plane* $\boldsymbol{\pi}$ *and the first image plane; and a perspectivity,* $\mathbf{x}' = \mathrm{H}_{2\pi}\mathbf{x}_{\pi}$, *between the world plane and second image plane. The composition of the two perspectivities is a homography,* $\mathbf{x}' = \mathrm{H}_{2\pi}\mathrm{H}_{1\pi}^{-1}\mathbf{x} = \mathrm{H}\mathbf{x}$, *between the image planes.*

There are then two relations between the two views: first, through the epipolar geometry a point in one view determines a line in the other which is the image of the ray through that point; and second, through the homography a point in one view determines a point in the other which is the image of the intersection of the ray with a plane. This chapter ties together these two relations of 2-view geometry.

Two other important notions are described here: the parallax with respect to a plane, and the infinite homography.

13.1 Homographies given the plane and vice versa

We start by showing that for planes in general position the homography is determined uniquely by the plane and vice versa. General position in this case means that the plane does not contain either of the camera centres. If the plane does contain one of the camera centres then the induced homography is degenerate.

Suppose a plane $\boldsymbol{\pi}$ in 3-space is specified by its coordinates in the world frame. We first derive an explicit expression for the induced homography.

Result 13.1. *Given the projection matrices for the two views*

$$\mathrm{P} = [\mathrm{I} \mid \mathbf{0}] \qquad \mathrm{P}' = [\mathrm{A} \mid \mathbf{a}]$$

and a plane defined by $\boldsymbol{\pi}^{\mathsf{T}}\mathbf{X} = 0$ *with* $\boldsymbol{\pi} = (\mathbf{v}^{\mathsf{T}}, 1)^{\mathsf{T}}$, *then the homography induced by the plane is* $\mathbf{x}' = \mathrm{H}\mathbf{x}$ *with*

$$\mathrm{H} = \mathrm{A} - \mathbf{a}\mathbf{v}^{\mathsf{T}}. \tag{13.1}$$

We may assume that $n_4 = 1$ since the plane does not pass through the centre of the first camera at $(0, 0, 0, 1)^T$.

Note, there is a three-parameter family of planes in 3-space, and correspondingly a three-parameter family of homographies between two views induced by planes in 3-space. These three parameters are specified by the elements of the vector **v**, which is *not* a homogeneous 3-vector.

Proof. To compute H we back-project a point **x** in the first view and determine the intersection point **X** of this ray with the plane ***π***. The 3D point **X** is then projected into the second view.
For the first view $\mathbf{x} = \mathrm{P}\mathbf{X} = [\mathrm{I} \mid \mathbf{0}]\mathbf{X}$ and so any point on the ray $\mathbf{X} = (\mathbf{x}^T, \rho)^T$ projects to **x**, where ρ parametrizes the point on the ray. Since the 3D point **X** is on ***π*** it satisfies $\boldsymbol{\pi}^T\mathbf{X} = 0$. This determines ρ, and $\mathbf{X} = (\mathbf{x}^T, -\mathbf{v}^T\mathbf{x})^T$. The 3D point **X** projects into the second view as

$$\begin{aligned} \mathbf{x}' &= \mathrm{P}'\mathbf{X} = [\mathrm{A} \mid \mathbf{a}]\mathbf{X} \\ &= \mathrm{A}\mathbf{x} - \mathbf{a}\mathbf{v}^T\mathbf{x} = \left(\mathrm{A} - \mathbf{a}\mathbf{v}^T\right)\mathbf{x} \end{aligned}$$

as required.

□

Example 13.2. A calibrated stereo rig.

Suppose the camera matrices are those of a calibrated stereo rig with the world origin at the first camera

$$\mathrm{P}_E = \mathrm{K}[\mathrm{I} \mid \mathbf{0}] \qquad \mathrm{P}'_E = \mathrm{K}'[\mathrm{R} \mid \mathbf{t}],$$

and the world plane $\boldsymbol{\pi}_E$ has coordinates $\boldsymbol{\pi}_E = (\mathbf{n}^T, d)^T$ so that for points on the plane $\mathbf{n}^T\tilde{\mathbf{X}} + d = 0$. We wish to compute an expression for the homography induced by the plane.

From result 13.1, with $\mathbf{v} = \mathbf{n}/d$, the homography for the cameras $\mathrm{P} = [\mathrm{I} \mid \mathbf{0}]$, $\mathrm{P}' = [\mathrm{R} \mid \mathbf{t}]$ is

$$H = R - \mathbf{t}\mathbf{n}^{\mathsf{T}}/d.$$

Applying the transformations K and K′ to the images we obtain the cameras $P_E = K[I \mid 0]$, $P'_E = K'[R \mid \mathbf{t}]$ and the resulting induced homography is

$$H = K' \left(R - \mathbf{t}\mathbf{n}^{\mathsf{T}}/d \right) K^{-1}. \tag{13.2}$$

This is a three-parameter family of homographies, parametrized by **n**/*d*. It is defined by the plane, and the camera internal and relative external parameters.

△

13.1.1 Homographies compatible with epipolar geometry

Suppose four points $\mathbf{X}_i$ are chosen on a scene plane. Then the correspondence $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ of their images between two views defines a homography H, which is the homography induced by the plane. These image correspondences also obey the epipolar constraint, i.e. $\mathbf{x}'^{\mathsf{T}}_i F \mathbf{x}_i = 0$, since they arise from images of scene points. Indeed, the correspondence $\mathbf{x} \leftrightarrow \mathbf{x}' = H\mathbf{x}$ obeys the epipolar constraint for *any* **x**, since again **x** and **x′** are images of a scene point, in this case the point given by intersecting the scene plane with the ray back-projected from **x.** The homography H is said to be consistent or *compatible* with F.

Now suppose four *arbitrary* image points are chosen in the first view and four arbitrary image points chosen in the second. Then a homography $\tilde{H}$ may be computed which maps one set of points into the other (provided no three are collinear in either view). However, correspondences $\mathbf{x} \leftrightarrow \mathbf{x}' = \tilde{H}\mathbf{x}$ may *not* obey the epipolar constraint. If the correspondence $\mathbf{x} \leftrightarrow \mathbf{x}' = \tilde{H}\mathbf{x}$ does not obey the epipolar constraint then there does not exist a scene plane which induces $\tilde{H}$.

The epipolar geometry determines the projective geometry between two views, and can be used to define conditions on homographies which are induced by actual scene planes. Figure 13.2 illustrates several relations between epipolar geometry and scene planes which can be used to define such conditions. For example, since correspondences **x** ↔ H**x** obey the epipolar constraint if H is induced by a plane, then from $\mathbf{x}'^{\mathsf{T}}\mathrm{F}\mathbf{x} = 0$

$$(\mathrm{H}\mathbf{x})^{\mathsf{T}}\mathrm{F}\mathbf{x} = \mathbf{x}^{\mathsf{T}}\mathrm{H}^{\mathsf{T}}\mathrm{F}\mathbf{x} = 0.$$

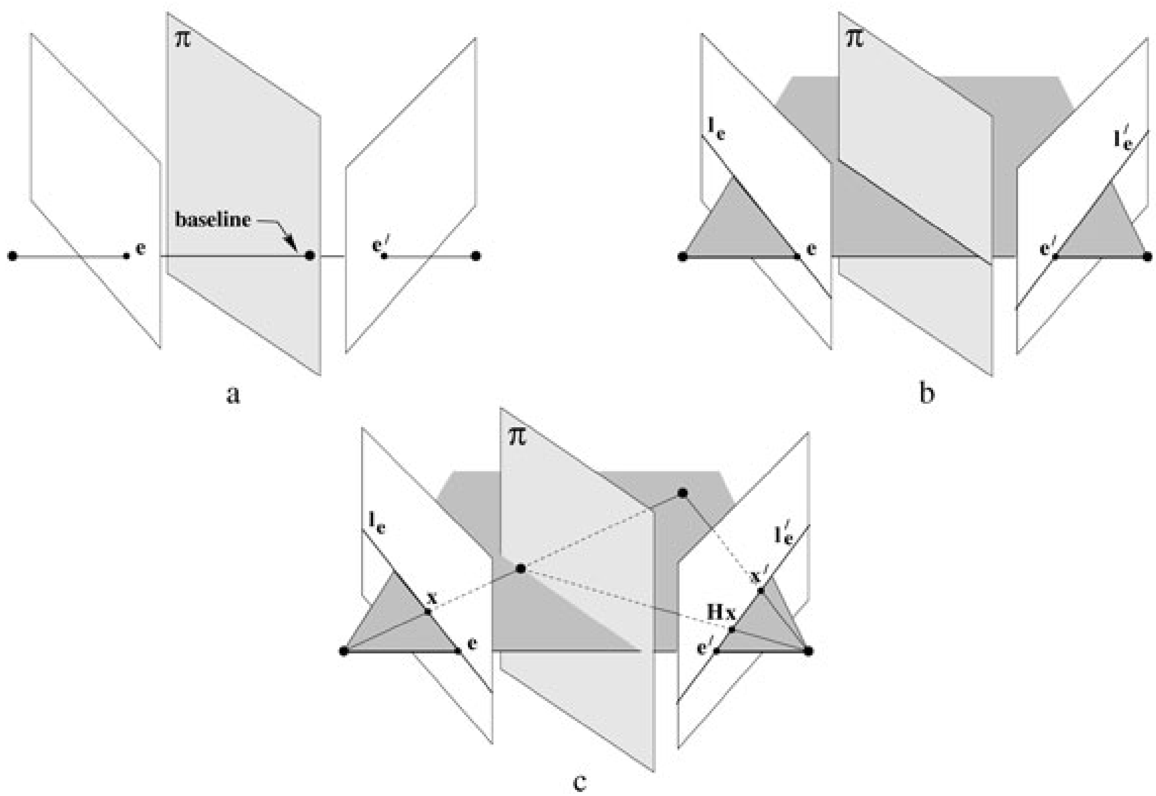


Fig. 13.2. **Compatibility constraints.** *The homography induced by a plane is coupled to the epipolar geometry and satisfies constraints. (a) The epipole is mapped by the homography, as* **e′** = H**e**, *since the epipoles are images of the point on the plane where the baseline intersects* ***π***. *(b) Epipolar lines are mapped by the homography as* $\mathrm{H}^{\mathsf{T}}\mathbf{l}'_e = \mathbf{l}_e$. *(c) Any point* **x** *mapped by the homography lies on its corresponding epipolar line* $\mathbf{l}'_e$, *so* $\mathbf{l}'_e = \mathrm{F}\mathbf{x} = \mathbf{x}' \times (\mathrm{H}\mathbf{x})$.

This is true for all **x**, so:

- *A homography* H *is compatible with a fundamental matrix* F *if and only if the matrix* $H^T F$ *is skew-symmetric:*

$$H^T F + F^T H = 0 \tag{13.3}$$

The argument above showed that the condition was necessary. The fact that this is a sufficient condition was shown by Luong and Viéville [Luong-96]. Counting degrees of freedom, (13.3) places six homogeneous (five inhomogeneous) constraints on the 8 degrees of freedom of H. There are therefore 8 − 5 = 3 degrees of freedom remaining for H; these 3 degrees of freedom correspond to the three-parameter family of planes in 3-space.

The compatibility constraint (13.3) is an implicit equation in H and F. We now develop an explicit expression for a homography H induced by a plane given F which is more suitable for a computational algorithm.

Result 13.3. *Given the fundamental matrix* F *between two views, the three-parameter family of homographies induced by a world plane is*

$$H = A - e' v^T \tag{13.4}$$

where $[\mathbf{e}']_\times A = F$ *is any decomposition of the fundamental matrix.*

Proof. Result 13.1 has shown that given the camera matrices for the view pair $P = [I \mid \mathbf{0}]$, $P' = [A \mid \mathbf{a}]$ a plane $\boldsymbol{\pi}$ induces a homography $H = A - \mathbf{a}\mathbf{v}^T$ where $\boldsymbol{\pi} = (\mathbf{v}^T, 1)^T$. However, according to result 9.9(*p*254), for the fundamental matrix $F = [\mathbf{e}']_\times A$ one can choose the two cameras to be $[I \mid \mathbf{0}]$ and $[A \mid \mathbf{e}']$.

□

Remark. The above derivation, which is based on the projection of points on a plane, ensures that the homographies are compatible with the epipolar geometry. Algebraically, the homography (13.4) is compatible with the fundamental matrix since it obeys the necessary and sufficient condition (13.3) that F^TH is skew-symmetric. This follows from

$$F^T H = A^T [e']_\times \left(A - e' v^T\right) = A^T [e']_\times A$$

using $[e']_\times e' = 0$, since $A^T[e']_\times A$ is skew-symmetric.

Comparing (13.4) with the general decomposition of the fundamental matrix, as given in lemma 9.11(*p*255) or (9.10–*p*256) it is evident that they involve an identical formula (except for signs). In fact there is a one-to-one correspondence between decompositions of the fundamental matrix (up to the scale factor ambiguity *k* in lemma 9.11) and homographies induced by world planes, as stated here.

Corollary 13.4. *A transformation* H *is the homography between two images induced by some world plane if and only if the fundamental matrix* F *for the two images has a decomposition* $F = [\mathbf{e}']_\times H$.

This choice in the decomposition simply corresponds to the choice of projective world frame. In fact, H is the transformation with respect to the plane with coordinates $(0, 0, 0, 1)^T$ in the reconstruction with $P = [I \mid \mathbf{0}]$ and $P' = [H \mid \mathbf{e}']$.

Finding the plane that induces a given homography is a simple matter given a pair of camera matrices, as follows.

Result 13.5. *Given the cameras in the canonical form* $P = [I \mid \mathbf{0}]$, $P' = [A \mid \mathbf{a}]$, *then the plane* $\boldsymbol{\pi}$ *that induces a given homography* H *between the views has coordinates* $\boldsymbol{\pi} = (\mathbf{v}^T, 1)^T$ *where* $\mathbf{v}$ *may be obtained linearly by solving the equations* $\lambda H = A - \mathbf{a}\mathbf{v}^T$, *which are linear in the entries of* $\mathbf{v}$ *and* λ.

Note, these equations have an exact solution only if H satisfies the compatibility constraint (13.3) with F. For a homography computed numerically from noisy data this will not normally be true, and the linear system is over-determined.

13.2 Plane induced homographies given F and image correspondences

A plane in 3-space can be specified by three points, or by a line and a point, and so forth. In turn these 3D elements can be specified by image correspondences. In section 13.1 the homography was computed from the coordinates of the plane. In the following the homography will be computed directly from the corresponding image elements that specify the plane. This is a quite natural mechanism to use in applications.

We will consider two cases: (i) three points; (ii) a line and a point. In each case the corresponding elements are sufficient to determine a plane in 3-space uniquely. It will be seen that in each case:

(i) The corresponding image entities have to satisfy *consistency constraints* with the epipolar geometry.

(ii) There are *degenerate configurations* of the 3D elements and cameras for which the homography is not defined. Such degeneracies arise from collinearities and coplanarities of the 3D elements and the epipolar geometry. There may also be degeneracies of the solution method, but these can be avoided.

The three-point case is covered in more detail.

13.2.1 Three points

We suppose that we have the images in two views of three (non-collinear) points $\mathbf{X}_i$, and the fundamental matrix F. The homography H induced by the plane of the points may be computed in principle in two ways:

First, the position of the points $\mathbf{X}_i$ is recovered in a projective reconstruction (chapter 12). Then the plane $\boldsymbol{\pi}$ through the points is determined (3.3–*p*66), and the homography computed from the plane as in result 13.1. Second, the homography may be computed from four corresponding points, the four points in this case being the images of the three points $\mathbf{X}_i$ on the plane together with the epipole in each view. The epipole may be used as the fourth point since it is mapped between the views by the homography as shown in figure 13.2. Thus we have four correspondences, $\mathbf{x}'_i = \mathtt{H}\mathbf{x}_i, i \in \{1, \ldots, 3\},\ \mathbf{e}' = \mathtt{H}\mathbf{e}$, from which H may be computed.

We thus have two alternative methods to compute H from three point correspondences, the first involving an *explicit* reconstruction, the second an *implicit* one where the epipole provides a point correspondence. It is natural to ask if one has an advantage over the other, and the answer is that the implicit method should **not** be used for computation as it has significant degeneracies which are not present in the explicit method.

Consider the case when two of the image points are collinear with the epipole (we assume for the moment that the measurements are noise-free). A homography H cannot be computed from four correspondences if three of the points are collinear (see section 4.1.3(*p*91)), so the implicit method fails in this case. Similarly if the image points are close to collinear with the epipole then the implicit method will give a poorly conditioned estimate for H. The explicit method has no problems when two points are collinear or close to collinear with the epipole – the corresponding image points define points in 3-space (the world points are on the same epipolar plane, but this is not a degenerate situation) and the plane ***π*** and hence homography can be computed. The configuration is illustrated in figure 13.3.

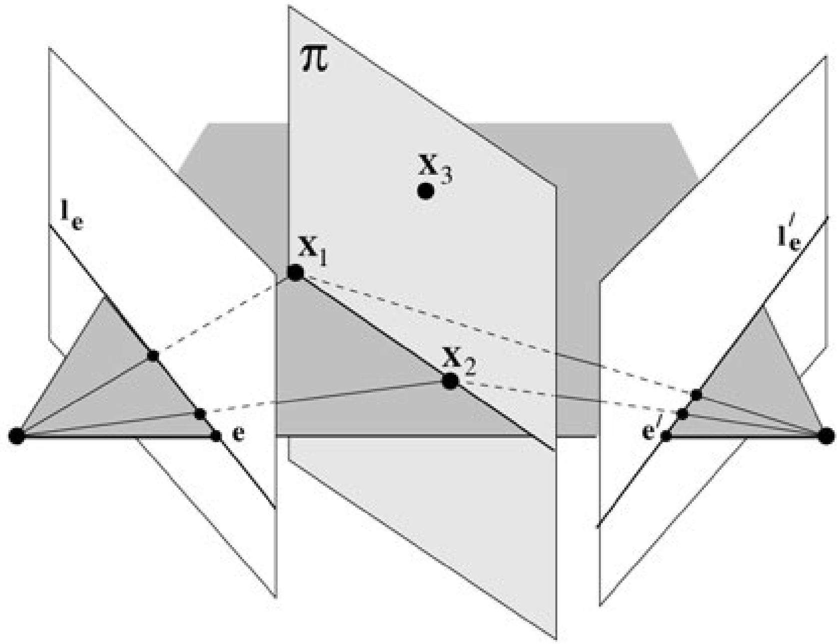


Fig. 13.3. **Degenerate geometry for an implicit computation of the homography.** *The line defined by the points* $\mathbf{X}_1$ *and* $\mathbf{X}_2$ *lies in an epipolar plane, and thus intersects the baseline. The images of* $\mathbf{X}_1$ *and* $\mathbf{X}_2$ *are collinear with the epipole, and* H *cannot be computed uniquely from the correspondences* $\mathrm{x}_i \leftrightarrow \mathrm{x}'_i, i \in \{1, \ldots, 3\}$, $\mathrm{e} \leftrightarrow \mathrm{e}'$. *This configuration is not degenerate for the explicit method.*

We now develop the algebra of the explicit method in more detail. It is not necessary to actually determine the coordinates of the points $\mathbf{X}_i$, all that is important is the constraint they place on the three-parameter family of homographies compatible with F (13.4), H = A − $\mathbf{e}'\mathbf{v}^\mathsf{T}$, parametrized by **v**. The problem is reduced to that of solving for **v** from the three point correspondences. The solution may be obtained as:

Result 13.6. *Given* F *and the three image point correspondences* $\mathrm{x}_i \leftrightarrow \mathrm{x}'_i$, *the homography induced by the plane of the 3D points is*

$$\mathtt{H} = \mathtt{A} - \mathrm{e}'(\mathtt{M}^{-1}\mathrm{b})^\mathsf{T},$$

where A = $[\mathbf{e}']_\times$F *and* **b** *is a 3-vector with components*

$$b_i = (\mathbf{x}'_i \times (\mathtt{A}\mathbf{x}_i))^\mathsf{T}(\mathbf{x}'_i \times \mathbf{e}')/\|\mathbf{x}'_i \times \mathbf{e}'\|^2,$$

and M *is a* 3 × 3 *matrix with rows* $\mathbf{x}_i^\mathsf{T}$.

Proof. According to result 9.14(*p*256), F may be decomposed as F = $[\mathbf{e}']_\times$A. Then (13.4) gives H = A − $\mathbf{e}'\mathbf{v}^\mathsf{T}$, and each correspondence $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ generates a linear constraint on **v** as

$$\mathbf{x}'_i = \mathtt{H}\mathbf{x}_i = \mathtt{A}\mathbf{x}_i - \mathbf{e}'(\mathbf{v}^\mathsf{T}\mathbf{x}_i), \quad i = 1, \ldots, 3. \tag{13.5}$$

From (13.5) the vectors $\mathbf{x}'_i$ and A$\mathbf{x}_i$ − $\mathbf{e}'(\mathbf{v}^\mathsf{T}\mathbf{x}_i)$ are parallel, so their vector product is zero:

$$\mathbf{x}'_i \times (\mathtt{A}\mathbf{x}_i - \mathbf{e}'(\mathbf{v}^\mathsf{T}\mathbf{x}_i)) = (\mathbf{x}'_i \times \mathtt{A}\mathbf{x}_i) - (\mathbf{x}'_i \times \mathbf{e}')(\mathbf{v}^\mathsf{T}\mathbf{x}_i) = \mathbf{0}.$$

Forming the scalar product with the vector $\mathbf{x}'_i$ × **e**′ gives

$$\mathbf{x}_i^\mathsf{T}\mathbf{v} = \frac{(\mathbf{x}'_i \times (\mathtt{A}\mathbf{x}_i))^\mathsf{T}(\mathbf{x}'_i \times \mathbf{e}')}{(\mathbf{x}'_i \times \mathbf{e}')^\mathsf{T}(\mathbf{x}'_i \times \mathbf{e}')} = b_i \tag{13.6}$$

which is linear in **v.** Note, the equation is independent of the scale of **x**′, since **x**′ occurs the same number of times in the numerator and denominator. Each correspondence generates an equation $\mathbf{x}_i^\mathsf{T}\mathbf{v} = b_i$, and collecting these together we have M**v** = **b**.

□

Note, a solution cannot be obtained if $\mathtt{M}^\mathsf{T} = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3]$ is not of full rank. Algebraically, det M = 0 if the three image points $\mathbf{x}_i$ are collinear. Geometrically, three collinear image points arise from collinear world points, or coplanar world points where the plane contains the first camera centre. In either case a full rank homography is not defined.

Consistency conditions. Equation (13.5) is equivalent to six constraints since each point correspondence places two constraints on a homography. Determining **v** requires only three constraints, so there are three constraints remaining which must be satisfied for a valid solution. These constraints are obtained by taking the cross product of (13.5) with **e**′, which gives

$$\mathbf{e}' \times \mathbf{x}'_i = \mathbf{e}' \times A\mathbf{x}_i = F\mathbf{x}_i.$$

Objective

Given F and three point correspondences $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ which are the images of 3D points $\mathbf{X}_i$, determine the homography $\mathbf{x}' = H\mathbf{x}$ induced by the plane of $\mathbf{X}_i$.

Algorithm

(i) For each correspondence $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ compute the corrected correspondence $\hat{\mathbf{x}}_i \leftrightarrow \hat{\mathbf{x}}'_i$ using algorithm 12.1(*p*318).

(ii) Choose $A = [\mathbf{e}']_\times F$ and solve linearly for **v** from $M\mathbf{v} = \mathbf{b}$ as in result 13.6.

(iii) Then $H = A - \mathbf{e}'\mathbf{v}^T$.

Algorithm 13.1. *The optimal estimate of the homography induced by a plane defined by three points.*

The equation $\mathbf{e}' \times \mathbf{x}'_i = F\mathbf{x}_i$ is a *consistency constraint* between $\mathbf{x}_i$ and $\mathbf{x}'_i$, since it is independent of **v**. It is simply a (disguised) epipolar constraint on the correspondence $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$: the LHS is the epipolar line through $\mathbf{x}'_i$, and the RHS is $F\mathbf{x}_i$ which is the epipolar line for $\mathbf{x}_i$ in the second image, i.e. the equation enforces that $\mathbf{x}'_i$ lie on the epipolar line of $\mathbf{x}_i$, and hence the correspondence is consistent with the epipolar geometry.

Estimation from noisy points. The three point correspondences which determine the plane and homography must satisfy the consistency constraint arising from the epipolar geometry. Generally *measured* correspondences $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ will not exactly satisfy this constraint. We therefore require a procedure for optimally correcting the measured points so that the estimated points $\hat{\mathbf{x}}_i \leftrightarrow \hat{\mathbf{x}}'_i$ satisfy the epipolar constraint. Fortunately, such a procedure has already been given in the triangulation algorithm 12.1(*p*318), which can be adopted here directly. We then have a Maximum Likelihood estimate of H and the 3D points under Gaussian image noise assumptions. The method is summarized in algorithm 13.1.

13.2.2 A point and line

In this section an expression is derived for a plane specified by a point and line correspondence. We start by considering only the line correspondence and show that this reduces the three-parameter family of homographies compatible with F (13.4) to a 1-parameter family. It is then shown that the point correspondence uniquely determines the plane and corresponding homography.

The correspondence of two image lines determines a line in 3-space, and a line in 3-space lies on a one parameter family (a pencil) of planes, see figure 13.4. This pencil of planes induces a pencil of homographies between the two images, and any member of this family will map the corresponding lines to each other.

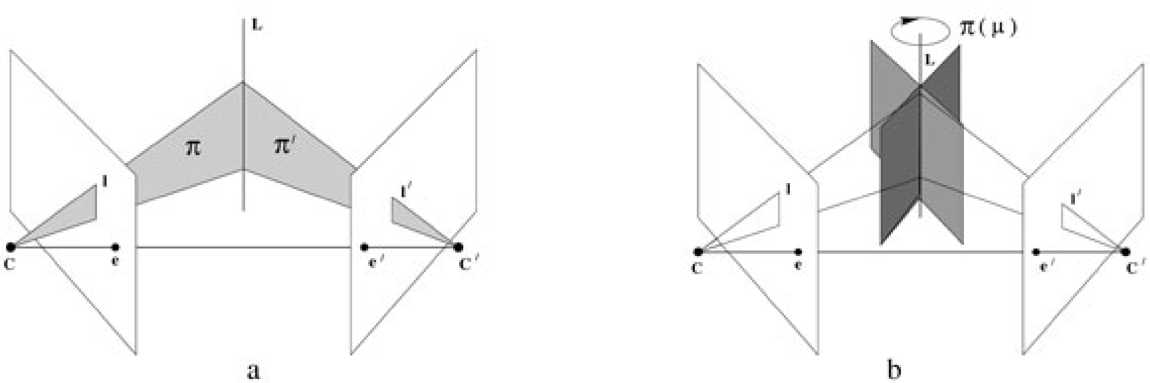


Fig. 13.4. *(a) Image lines* **l** *and* **l'** *determine planes* ***π*** *and* ***π'*** *respectively. The intersection of these planes defines the line* **L** *in 3-space. (b) The line* **L** *in 3-*

space is contained in a one parameter family of planes $\boldsymbol{\pi}(\mu)$. *This family of planes induces a one parameter family of homographies between the images.*

Result 13.7. *The homography for the pencil of planes defined by a line correspondence* $\mathbf{l} \leftrightarrow \mathbf{l}'$ *is given by*

$$H(\mu) = [l']_\times F + \mu e' l^T \tag{13.7}$$

provided $\mathbf{l}'^T\mathbf{e}' \neq 0$, *where* μ *is a projective parameter.*

Proof. From result 8.2(*p*197) the line $\mathbf{l}$ back-projects to a plane $P^T\mathbf{l}$ through the first camera centre, and $\mathbf{l}'$ back-projects to a plane $P'^T\mathbf{l}'$ through the second, see figure 13.4a. These two planes are the basis for a pencil of planes parametrized by μ. As in the proof of result 13.3 we may choose $P = [I \mid \mathbf{0}]$, $P' = [A \mid \mathbf{e}']$, then the pencil of planes is

$$\begin{aligned} \boldsymbol{\pi}(\mu) &= \mu P^T l + P'^T l' \\ &= \mu \begin{pmatrix} l \\ 0 \end{pmatrix} + \begin{pmatrix} A^T l' \\ e'^T l' \end{pmatrix} \end{aligned}$$

From result 13.1 the induced homography is $H(\mu) = A - \mathbf{e}'\mathbf{v}(\mu)^T$, with

$$v(\mu) = (\mu l + A^T l')/(e'^T l') \tag{13.8}$$

Using the decomposition $A = [\mathbf{e}']_\times F$ we obtain

$$\begin{aligned} H &= \left((e'^T l' I - e' l'^T)[e']_\times F - \mu e' l^T\right)/(e'^T l') = -\left([l']_\times [e']_\times [e']_\times F + \mu e' l^T\right)/(e'^T l') \\ &= -\left([l']_\times F + \mu e' l^T\right)/(e'^T l') \end{aligned}$$

where the last equality follows from result A4.4(*p*582) that $[\mathbf{e}']_\times[\mathbf{e}']_\times F = F$. This is equivalent to (13.7) up to scale.

□

The homography for a corresponding point and line. From the line correspondence we have that $H(\mu) = [\mathbf{l}']_\times F + \mu \mathbf{e}' \mathbf{l}^\mathsf{T}$, and now solve for μ using the point correspondence $\mathbf{x} \leftrightarrow \mathbf{x}'$.

Result 13.8. *Given* F *and a corresponding point* $\mathbf{x} \leftrightarrow \mathbf{x}'$ *and line* $\mathbf{l} \leftrightarrow \mathbf{l}'$, *the homography induced by the plane of the 3-space point and line is*

$$H = [\mathbf{l}']_\times F + \frac{(\mathbf{x}' \times \mathbf{e}')^\mathsf{T} (\mathbf{x}' \times ((F\mathbf{x}) \times \mathbf{l}'))}{\|\mathbf{x}' \times \mathbf{e}'\|^2 (\mathbf{l}^\mathsf{T} \mathbf{x})} \mathbf{e}' \mathbf{l}^\mathsf{T}.$$

The derivation is analogous to that of result 13.6. As in the three-point case, the image point correspondence must be consistent with the epipolar geometry. This means that the measured (noisy) points must be corrected using algorithm 12.1(*p*318) before result 13.8 is applied. There is no consistency constraint on the line, and no correction is available.

Geometric interpretation of the point map $H(\mu)$. It is worth exploring the map $H(\mu)$ further. Since $H(\mu)$ is compatible with the epipolar geometry, a point $\mathbf{x}$ in the first view is mapped to a point $\mathbf{x}' = H(\mu)\mathbf{x}$ in the second view on the epipolar line $F\mathbf{x}$ corresponding to $\mathbf{x}$. In general the position of the point $\mathbf{x}' = H(\mu)\mathbf{x}$ on the epipolar line varies with μ. However, if the point $\mathbf{x}$ lies on $\mathbf{l}$ (so that $\mathbf{l}^\mathsf{T}\mathbf{x} = 0$) then

$$\mathbf{x}' = H(\mu)\mathbf{x} = ([\mathbf{l}']_\times F + \mu \mathbf{e}' \mathbf{l}^\mathsf{T})\mathbf{x} = [\mathbf{l}']_\times F\mathbf{x}$$

which is independent of the value of μ, depending only on F. Thus as shown in figure 13.5 the epipolar geometry defines a point-to-point map for points on the line.

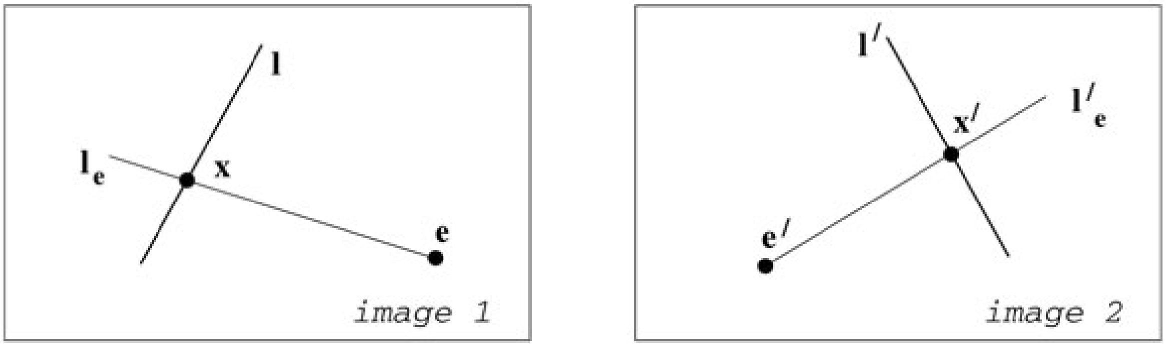


Fig. 13.5. *The epipolar geometry induces a homography between corresponding lines* **l** ↔ **l′** *which are the images of a line* **L** *in 3-space. The points on* **l** *are mapped to points on* **l′** *as* **x′** = [**l′**]$_\times$F**x**, *where* **x** *and* **x′** *are the images of the intersection of* **L** *with the epipolar plane corresponding to* $\mathbf{l}_e$ *and* $\mathbf{l}'_e$.

Degenerate homographies. As has already been stated, if the world plane contains one of the camera centres, then the induced homography is degenerate. The matrix representing the homography does not have full rank, and points on one plane are mapped to a line (if rank H = 2) or a point (if rank H = 1). However, an explicit expression can be obtained for a degenerate homography from (13.7). The degenerate (singular) homographies in this pencil are at $\mu = \infty$ and $\mu = 0$. These correspond to planes through the first and second camera centres respectively. Figure 13.6 shows the case where the plane contains the second camera centre, and intersects the image plane in the line **l′**. A point **x** in the first view is imaged on **l′** at the point **x′** where

$$\mathbf{x}' = \mathbf{l}' \times \mathrm{F}\mathbf{x} = [\mathbf{l}']_\times \mathrm{F}\mathbf{x}.$$

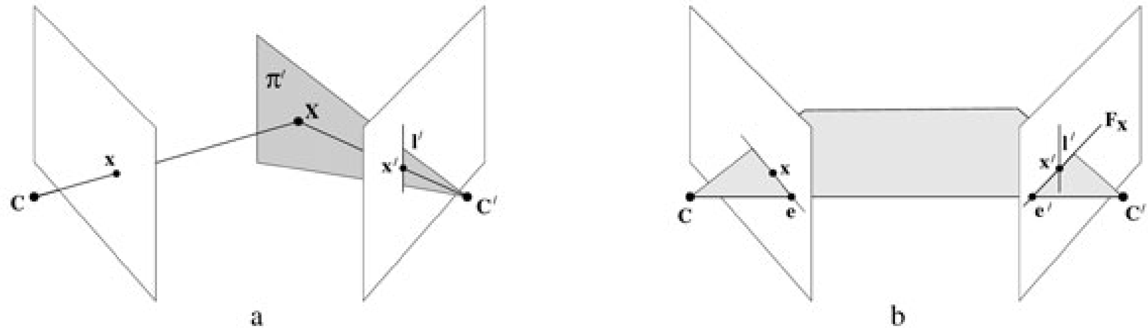


Fig. 13.6. **A degenerate homography.** *(a) The map induced by a plane through the second camera centre is a degenerate homography* $H = [\mathbf{l}']_\times F$. *The plane* $\boldsymbol{\pi}'$ *intersects the second image plane in the line* $\mathbf{l}'$. *All points in the first view are mapped to points on* $\mathbf{l}'$ *in the second. (b) A point* $\mathbf{x}$ *in the first view is imaged at* $\mathbf{x}'$, *the intersection of* $\mathbf{l}'$ *with the epipolar line* $F\mathbf{x}$ *of* $\mathbf{x}$, *so that* $\mathbf{x}' = \mathbf{l}' \times F\mathbf{x}$.

The homography is thus $H = [\mathbf{l}']_\times F$. This is a rank 2 matrix.

13.3 Computing F given the homography induced by a plane

Up to now it has been assumed that F is given, and the objective is to compute H when various additional information is provided. We now reverse this, and show that if H is given then F may be computed when additional information is provided. We start by introducing an important geometric idea, that of parallax relative to a plane, which will make the algebraic development straightforward.

Plane induced parallax. The homography induced by a plane generates a virtual parallax (see section 8.4.5(*p*207)) as illustrated schematically in figure 13.7 and by example in figure 13.8. The important point here is that in the second view $\mathbf{x}'$, the image of the 3D point $\mathbf{X}$, and $\tilde{\mathbf{x}}' = H\mathbf{x}$, the point mapped by the homography, are on the epipolar line of $\mathbf{x}$; since both are images of points on the ray through $\mathbf{x}$. Consequently, the line $\mathbf{x}' \times (H\mathbf{x})$ is an epipolar line in the second view and provides a constraint on the position of the epipole. Once the epipole is determined (two such constraints suffice), then as shown in result 9.1(*p*243) F = $[\mathbf{e}']_\times$H where H is the homography induced by any plane. Similarly it can be shown that F = $H^{-T}[\mathbf{e}]_\times$.

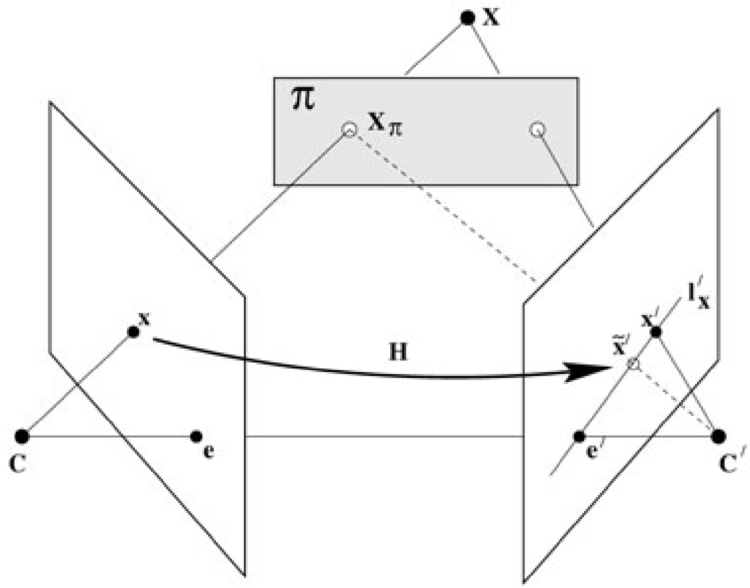


Fig. 13.7. **Plane induced parallax.** *The ray through* **X** *intersects the plane* ***π*** *at the point* $\mathbf{X}_\pi$. *The images of* **X** *and* $\mathbf{X}_\pi$ *are coincident points at* **x** *in the first view. In the second view the images are the points* **x′** *and* $\tilde{\mathbf{x}}'$ = H**x** *respectively. These points are not coincident (unless* **X** *is on* ***π****), but both are on the epipolar line* $\mathbf{l}'_\mathbf{x}$ *of* **x**. *The vector between the points* **x′** *and* $\tilde{\mathbf{x}}'$ *is the* parallax relative to the homography *induced by the plane* ***π***. *Note that if* **X** *is on the other side of the plane, then* $\tilde{\mathbf{x}}'$ *will be on the other side of* **x′**.

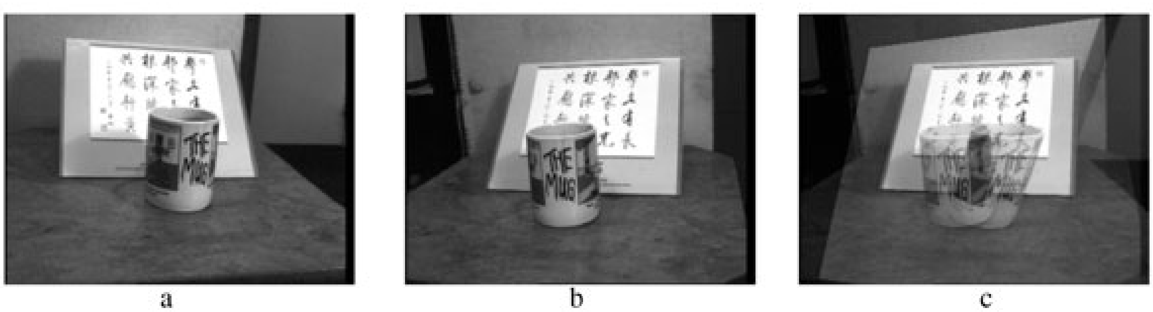


Fig. 13.8. **Plane induced parallax.** *(a) (b) Left and right images. (c) The left image is superimposed on the right using the homography induced by the plane of the Chinese text. The transferred and imaged planes exactly coincide. However, points off the plane (such as the mug) do not coincide. Lines joining*

corresponding points off the plane in the "superimposed" image intersect at the epipole.

As an application of virtual parallax it is shown in algorithm 13.2 that F can be computed uniquely from the images of six points, four of which are coplanar and two are off the plane. The images of the four coplanar points define the homography, and the two points off the plane provide constraints sufficient to determine the epipole. The six-point result is quite surprising since for seven points in general position there are 3 solutions for F (see section 11.1.2(*p*281)).

Objective

Objective Given six point correspondences $x_i \leftrightarrow x'_i$ which are the images of 3-space $\mathbf{X}_i$, with the first four 3-space points $i \in \{1, \ldots, 4\}$ coplanar, determine the fundamental matrix F.

Algorithm

(i) Compute the homography H, such that $x'_i = Hx_i, i \in \{1, \ldots, 4\}$.

(ii) Determine the epipole $\mathbf{e}'$ as the intersection of the lines $(Hx_5) \times x'_5$ and $(Hx_6) \times x'_6$.

(iii) Then $F = [\mathbf{e}']_\times H$.

See figure 13.9.

Algorithm 13.2. *Computing* F *given the correspondences of six points of which four are coplanar.*

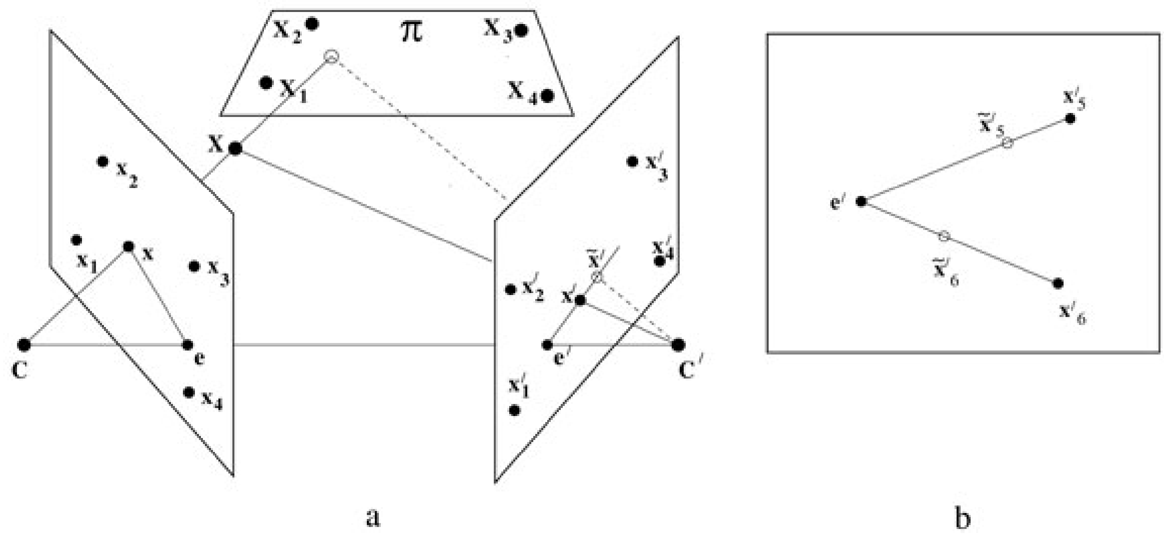


Fig. 13.9. *The fundamental matrix is defined uniquely by the image of six 3D points, of which four are coplanar. (a) The parallax for one point* **X**. *(b) The epipole determined by the intersection of two parallax lines: the line joining* $\tilde{x}'_5 = Hx_5$ *to* x'_5, *and the join of* $\tilde{x}'_6 = Hx_6$ *to* x'_6.

Projective depth. A world point $\mathbf{X} = (\mathbf{x}^T, \rho)^T$ is imaged at $\mathbf{x}$ in the first view and at

$$x' = Hx + \rho e' \tag{13.9}$$

in the second. Note that $\mathbf{x}'$, $\mathbf{e}'$ and H$\mathbf{x}$ are collinear. The scalar ρ is the parallax *relative* to the homography H, and may be interpreted as a "depth" relative to the plane ***π***. If $\rho = 0$ then the 3D point **X** is on the plane, otherwise the "sign" of ρ indicates which 'side' of the plane ***π*** the point **X** is (see figure 13.7 and figure 13.8). These statements should be taken with care because in the absence of oriented projective geometry the sign of a homogeneous object, and the side of a plane have no meaning.

Example 13.9. Binary space partition. The sign of the virtual parallax (sign(ρ)) may be used to compute a partition of 3-space by the plane ***π***. Suppose we are given F and three space points are specified by their corresponding image points. Then the plane defined by the three points can be used to partition all other correspondences into sets on either side of (or on) the plane. Figure 13.10 shows an example. Note, the three points need not actually correspond to images of physical points so the method can be applied to virtual planes. By combining several planes a region of 3-space can be identified.

△

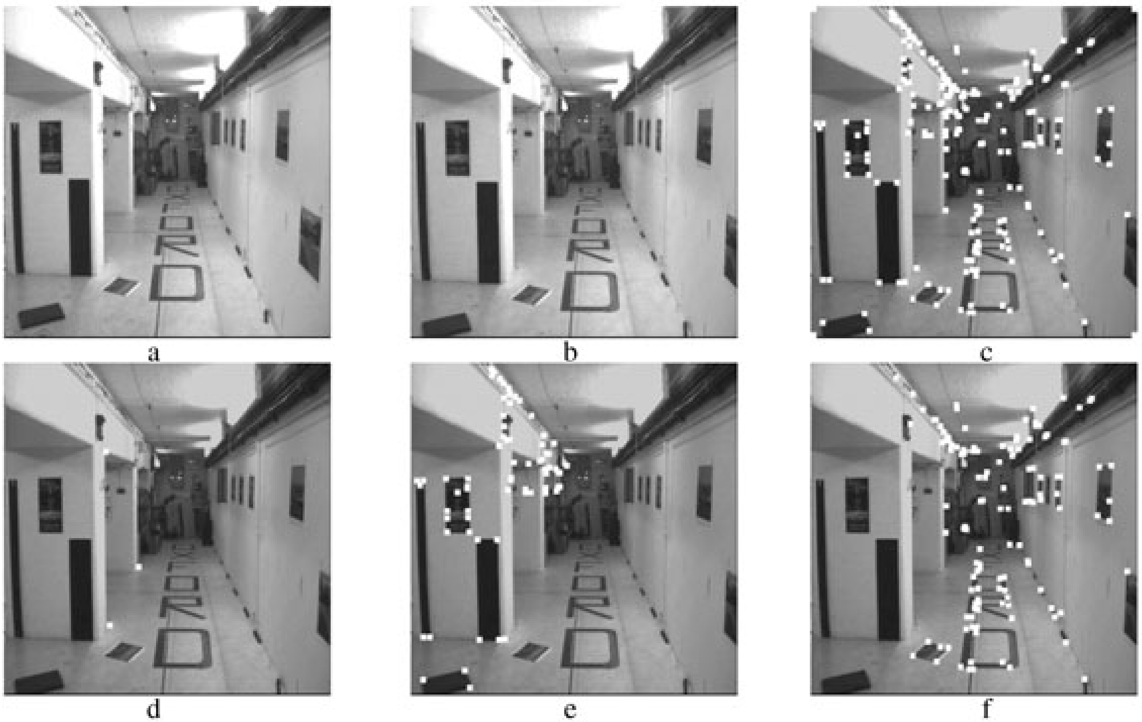


Fig. 13.10. **Binary space partition.** *(a) (b) Left and right images. (c) Points whose correspondence is known. (d) A triplet of points selected from (c). This triplet defines a plane. The points in (c) can then be classified according to their side of the plane. (e) Points on one side. (f) Points on the other side.*

Two planes. Suppose there are two planes, $\boldsymbol{\pi}_1$, $\boldsymbol{\pi}_2$, in the scene which induce homographies H_1, H_2 respectively. With the idea of parallax in mind it is clear that because each plane provides off-plane information about the other, the two homographies should be sufficient to determine F. Indeed F is over-determined by this configuration which means that the two homographies must satisfy consistency constraints.

Consider figure 13.11. The homography $H = H_2^{-1}H_1$ is a mapping from the first image onto itself. Under this mapping the epipole **e** is a fixed point, i.e. H**e** = **e**, so may be determined from the (non-degenerate) eigenvector of H. The fundamental matrix may then be computed from result 9.1(*p*243) as $F = [\mathbf{e}']_\times H_i$, where $\mathbf{e}' = H_i\mathbf{e}$ for i = 1 or 2. The map H has further properties which may be seen from figure 13.11. The map has a line of fixed points and a fixed point not on the line (see section 2.9(*p*61) for fixed points and lines). This means that two of the eigenvalues of H are equal. In fact H is a planar homology (see section A7.2(*p*629)). In turn, these properties of $H = H_2^{-1}H_1$ define consistency constraints on H_1 and H_2 in order that their composition has these properties.

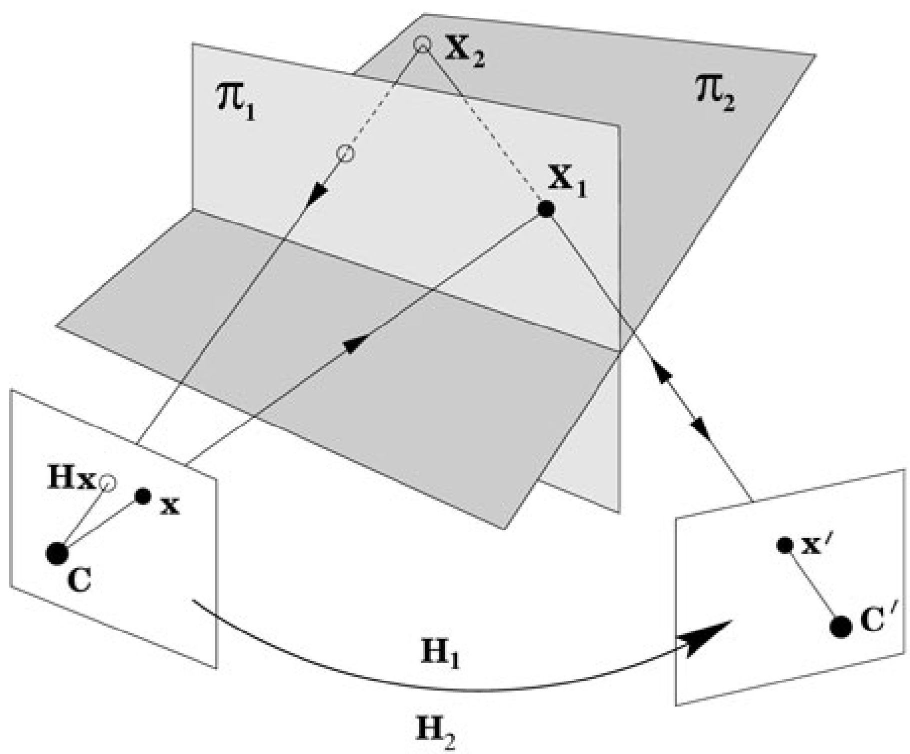


Fig. 13.11. *The action of the map* $H = H_2^{-1}H_1$ *on a point* **x** *in the first image is first to transfer it to* **x′** *as though it were the image of the 3D point* $\mathbf{X}_1$*, and then map it back to the first image as though it were the image of the 3D point* $\mathbf{X}_2$*. Points in the first view which lie on the imaged line of intersection of the two planes will be mapped to themselves, so are fixed points under this action. The epipole* **e** *is also a fixed point under this map.*

Up to this point the results of this chapter have been entirely projective. Now an affine element is introduced.

13.4 The infinite homography H_∞

The plane at infinity is a particularly important plane, and the homography induced by this plane is distinguished by a special name:

Definition 13.10. The infinite homography, H_∞, is the homography induced by the plane at infinity, $\boldsymbol{\pi}_\infty$.

The form of the homography may be derived by a limiting process starting from (13.2–*p*327), $H = K'(R - \mathbf{t}\mathbf{n}^\mathsf{T}/d)K^{-1}$, where d is the orthogonal distance of the plane from the first camera:

$$H_\infty = \lim_{d\to\infty} H = K'RK^{-1}.$$

This means that H_∞ does not depend on the translation between views, only on the rotation and camera internal parameters. Alternatively, from (9.7–*p*250) corresponding image points are related as

$$x' = K'RK^{-1}x + K't/Z = H_\infty x + K't/Z \tag{13.10}$$

where Z is the depth measured from the first camera. Again it can be seen that points at infinity ($Z = \infty$) are mapped by H_∞. Note also that H_∞ is obtained if the translation $\mathbf{t}$ is zero in (13.10), which corresponds to a rotation about the camera centre. Thus H_∞ is the homography that relates image points of *any* depth if the camera rotates about its centre (see section 8.4(*p*202)).

Since $\mathbf{e}' = K'\mathbf{t}$, (13.10) can be written as $\mathbf{x}' = H_\infty\mathbf{x} + \mathbf{e}'/Z$, and comparison with (13.9) shows that $(1/Z)$ plays the role of ρ. Thus Euclidean inverse depth can be interpreted as parallax relative to $\boldsymbol{\pi}_\infty$.

Vanishing points and lines. Images of points on π_∞ are mapped by H_∞. These images are vanishing points, and so H_∞ maps vanishing points between images, i.e. $\mathbf{v}' = H_\infty \mathbf{v}$, where $\mathbf{v}'$ and $\mathbf{v}$ are corresponding vanishing points. See figure 13.12. Consequently, H_∞ can be computed from the correspondence of three (non-collinear) vanishing points together with F using result 13.6. Alternatively, H_∞ can be computed from the correspondence of a vanishing line and the correspondence of a vanishing point (not on the line), together with F, as described in section 13.2.2.

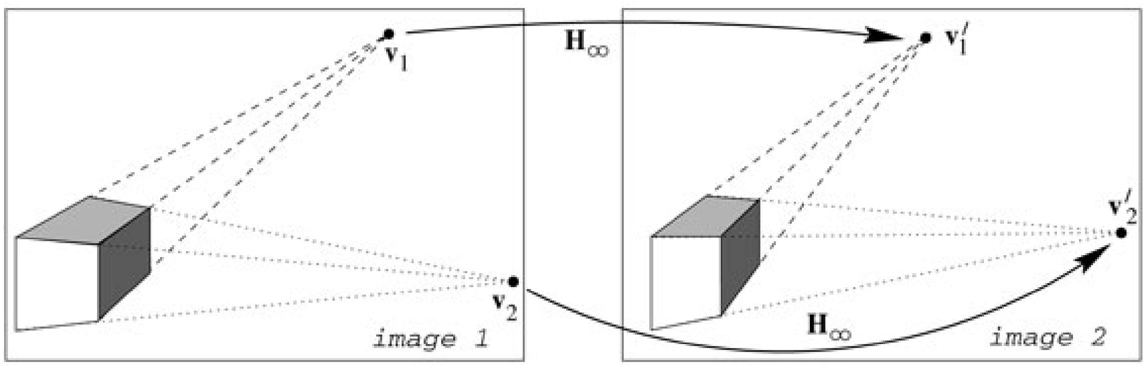


Fig. 13.12. *The infinite homography* H_∞ *maps vanishing points between the images.*

Affine and metric reconstruction. As we have seen in chapter 10, specifying π_∞ enables a projective reconstruction to be upgraded to an affine reconstruction. Not surprisingly, because of its association with π_∞, H_∞ arises naturally in the rectification. Indeed, if the camera matrices are chosen as $P = [I \mid \mathbf{0}]$ and $P' = [H_\infty \mid \lambda \mathbf{e}']$ then the reconstruction is affine.

Conversely, suppose the world coordinate system is affine (i.e. π_∞ has its canonical position at $\pi_\infty = (0, 0, 0, 1)^T$); then H_∞ may be determined directly from the camera projection matrices. Suppose M, M′ are the first 3 × 3 submatrix of P and P′ respectively. Then a point $\mathbf{X} = (\mathbf{x}_\infty^T, 0)^T$ on π_∞ is imaged at

$\mathbf{x} = \mathrm{P}\mathbf{X} = \mathrm{M}\mathbf{x}_\infty$ and $\mathbf{x}' = \mathrm{P}'\mathbf{X} = \mathrm{M}'\mathbf{x}_\infty$ in the two views. Consequently $\mathbf{x}' = \mathrm{M}'\mathrm{M}^{-1}\mathbf{x}$ and so

$$\mathrm{H}_\infty = \mathrm{M}'\mathrm{M}^{-1}. \tag{13.11}$$

The homography H_∞ may be used to propagate camera calibration from one view to another. The absolute conic Ω_∞ resides on $\boldsymbol{\pi}_\infty$, and its image, $\boldsymbol{\omega}$, is mapped between images by H_∞ according to result 2.13(*p*37): $\omega' = \mathrm{H}_\infty^{-\mathsf{T}}\omega\mathrm{H}_\infty^{-1}$. Thus if $\boldsymbol{\omega} = (\mathrm{K}\mathrm{K}^\mathsf{T})^{-1}$ is specified in one view, then $\boldsymbol{\omega}'$, the image of Ω_∞ in a second view, can be computed via H_∞, and the calibration for that view determined from $\boldsymbol{\omega}' = (\mathrm{K}'\mathrm{K}'^\mathsf{T})^{-1}$. Section 19.5.2(*p*475) describes applications of H_∞ to camera auto-calibration.

Stereo correspondence. H_∞ limits the search region when searching for correspondences. The region is reduced from the entire epipolar line to a bounded line. See figure 13.13. However, a correct application of this constraint requires oriented projective geometry.

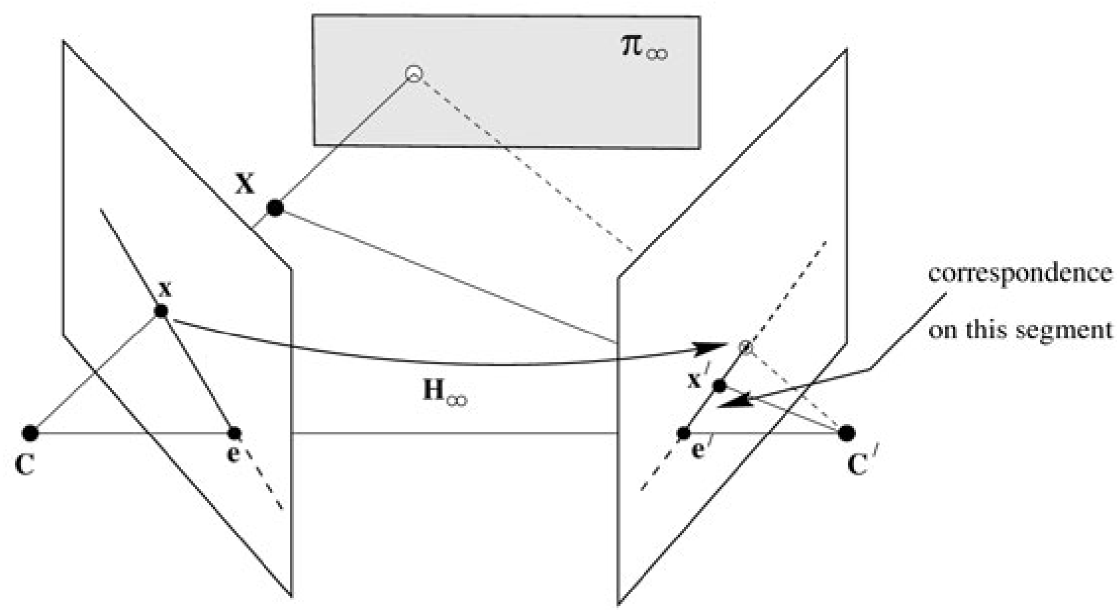


Fig. 13.13. **Reducing the search region using** H_∞. *Points in 3-space are no 'further' away than* $\boldsymbol{\pi}_\infty$. H_∞ *captures this constraint and limits the search on*

the epipolar line in one direction. The baseline between the cameras partitions each epipolar plane into two. A point on one "side" of the epipolar line in the left image will be imaged on the corresponding "side" of the epipolar line in the right image (indicated by the solid line in the figure). The epipole thus bounds the search region in the other direction.

13.5 Closure

This chapter has illustrated a raft of projective techniques for a plane that may be applied to many other surfaces. A plane is a simple parametrized surface with 3 degrees of freedom. A very similar development can be given for other surfaces where the degrees of freedom are determined from images of points on the surface. For example in the case of a quadric the surface can be determined both from images of points on its surface, and/or (an extension not possible for planes) from its outline in each view [Cross-98, Shashua-97]. The ideas of surface induced transfer, families of surfaces when the surface is not fully determined from its images, surface induced parallax, consistency constraints, implicit computations, degenerate geometries etc. all carry over to other surfaces.

13.5.1 The literature

The compatibility of epipolar geometry and induced homographies is investigated thoroughly by Luong & Viéville [Luong-96]. The six-point solution for F appeared in Beardsley *et al.* [Beardsley-92] and [Mohr-92]. The solution for F given two planes appeared in Sinclair [Sinclair-92]. [Zeller-96] gives many examples of configurations whose properties may be determined using only epipolar geometry and their image projections. He also catalogues their degenerate cases.

13.5.2 Notes and exercises

(i) **Homography induced by a plane (13.1–*p*326).**

(a) The inverse of the homography H is given by

$$H^{-1} = A^{-1}\left(I + \frac{\mathbf{a}\mathbf{v}^T A^{-1}}{1 - \mathbf{v}^T A^{-1}\mathbf{a}}\right)$$

provided A^{-1} exists. This is sometimes called the Sherman-Morrison formula.

(b) Show that the homography H is degenerate if the plane contains the second camera centre. Hint, in this case $\mathbf{v}^T A^{-1}\mathbf{a} = 1$, and note that $H = A(I - A^{-1}\mathbf{a}\mathbf{v}^T)$.

(ii) Show that if the camera undergoes planar motion, i.e. the translation is parallel to the plane and the rotation is parallel to the plane normal, then the homography induced by the plane is conjugate to a planar Euclidean transformation. Show that the fixed points of the homography are the images of the plane's circular points.

(iii) Using (13.2–*p*327) show that if a camera undergoes a pure translation then the homography induced by the plane is a planar homology (as defined in section A7.2(*p*629)), with a line of fixed points corresponding to the vanishing line of the plane. Show further that if the translation is parallel to the plane then the homography is an elation (as defined in section A7.3(*p*631)).

(iv) Show that a necessary, but not sufficient, condition for two space lines to be coplanar is $(\mathbf{l}'_1 \times \mathbf{l}'_2)^T F(\mathbf{l}_1 \times \mathbf{l}_2) = 0$. Why is it not a sufficient condition?

(v) **Intersections of lines and planes.** Verify each of the following results by sketching the configuration assuming general position. In each case determine the degenerate configurations for which the result is not valid.

(a) Suppose the line $\mathbf{L}$ in 3-space is imaged as $\mathbf{l}$ and $\mathbf{l}'$, and the plane $\boldsymbol{\pi}$ induces the homography $\mathbf{x}' = H\mathbf{x}$. Then the point of intersection of $\mathbf{L}$ with $\boldsymbol{\pi}$ is imaged at $\mathbf{x} = \mathbf{l} \times (H^T\mathbf{l}')$ in the first image, and at $\mathbf{x}' = \mathbf{l}' \times (H^{-T}\mathbf{l})$ in the second.

(b) The infinite homography may be used to find the vanishing point of a line seen in two images. If **l** and **l**′ are corresponding lines in two images, and **v**, **v**′ their vanishing points in each image, then **v** = **l** × $(\mathrm{H}_\infty^\mathsf{T}\mathbf{l}'),\ \mathbf{v}' = \mathbf{l}' \times (\mathrm{H}_\infty^{-\mathsf{T}}\mathbf{l})$.

(c) Suppose the planes π_1 and π_2 induce homographies **x**′ = H_1**x** and **x**′ = H_2**x** respectively. Then the image of the line of intersection of π_1 with π_2 in the first image obeys $\mathrm{H}_1^\mathsf{T}\mathrm{H}_2^{-\mathsf{T}}\mathbf{l} = \mathbf{l}$ and may be determined from the real eigenvector of the planar homology $\mathrm{H}_1^\mathsf{T}\mathrm{H}_2^{-\mathsf{T}}$ (see figure 13.11).

(vi) **Coplanarity of four points.** Suppose F is known, and four corresponding image points $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ are supplied. How can it be determined whether their pre-images are coplanar? One possibility is to use three of the points to determine a homography via result 13.6(*p*331), and then measure the transfer error of the fourth point. A second possibility is to compute lines joining the image points, and determine if the line intersection obeys the epipolar constraint (see [Faugeras-92b]). A third possibility is to compute the cross-ratio of the four lines from the epipole to the image points – if the four scene points are coplanar then this cross-ratio will be the same in both images. Thus this equality is a necessary condition for co-planarity, but is it a sufficient condition also? What statistical tests should be applied when there is measurement error (noise)?

(vii) Show that the epipolar geometry can be computed uniquely from the images of four coplanar lines and two points off the plane of the lines. If two of the lines are replaced by points can the epipolar geometry still be computed?

(viii) Starting from the camera matrices P = [M / **m**], P′ = [M′ / **m**′] show that the homography **x**′ = H**x** induced by a plane $\pi = (\tilde{\pi}^\mathsf{T}, \pi_4)^\mathsf{T}$ is given by

$$\mathrm{H} = \mathrm{M}'(\mathrm{I} - \mathbf{t}\mathbf{v}^\mathsf{T})\mathrm{M}^{-1} \text{ with } \mathbf{t} = (\mathrm{M}'^{-1}\mathbf{m}' - \mathrm{M}^{-1}\mathbf{m}), \text{ and } \mathbf{v} = \tilde{\pi}/(\pi_4 - \tilde{\pi}^\mathsf{T}\mathrm{M}^{-1}\mathbf{m}).$$

(ix) Show that the homography computed as in result 13.6(*p*331) is independent of the scale of F. Start by choosing an arbitrary fixed scale for F, so that F is no longer a homogeneous quantity, but a matrix $\tilde{F}$ with fixed scale. Show that if $H = [e']_\times \tilde{F} - e'(M^{-1}\tilde{b})^\mathsf{T}$ with $\tilde{b}_i = c_i'^\mathsf{T}(\tilde{F}x_i)$, then replacing $\tilde{F}$ by $\lambda\tilde{F}$ simply scales H by λ.

(x) Given two perspective images of a (plane) conic and the fundamental matrix between the views, then the plane of the conic (and consequently the homography induced by this plane) is defined up to a two-fold ambiguity. Suppose the image conics are C and C′, then the induced homography is $H(\mu) = [C'\mathbf{e}']_\times F - \mu \mathbf{e}'(C\mathbf{e})^\mathsf{T}$, with the two values of μ obtained from

$$\mu^2 \left[(e^\mathsf{T}Ce)C - (Ce)(Ce)^\mathsf{T}\right](e'^\mathsf{T}C'e') = -F^\mathsf{T}[C'e']_\times C'[C'e']_\times F.$$

Details are given in [Schmid-98].

(a) By considering the geometry, show that to be compatible with the epipolar geometry the conics must satisfy the consistency constraint that epipolar tangents are corresponding epipolar lines (see figure 11.6-(*p*295)). Now derive this result algebraically starting from $H(\mu)$ above.

(b) The algebraic expressions are not valid if the epipole lies on the conic (since then $\mathbf{e}^\mathsf{T}C\mathbf{e} = \mathbf{e}'^\mathsf{T}C'\mathbf{e}' = 0$). Is this a degeneracy of the geometry or of the expression alone?

(xi) **Fixed points of a homography induced by a plane.** A planar homography H has up to three distinct fixed points corresponding to the three eigenvectors of the 3 × 3 matrix (see section 2.9(*p*61)). The fixed points are images of points on the plane for which $\mathbf{x}' = H\mathbf{x} = \mathbf{x}$. The horopter is the locus of *all* points in 3-space for which $\mathbf{x} = \mathbf{x}'$. It is a twisted cubic curve passing through the two camera centres. A twisted cubic intersects a plane in three points, and these are the three fixed points of the homography induced by that plane.

(xii) **Estimation.** Suppose $n > 3$ points $\mathbf{X}_i$ lie on a plane in 3-space and we wish to optimally estimate the homography induced by the plane given F and their image correspondences $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$. Then the ML estimate of the homography (assuming independent Gaussian measurement noise as usual) is obtained by estimating the plane $\hat{\pi}$ (3 dof) and the n points $\hat{\mathbf{X}}_i$ (2 dof each, since they lie on a plane) which minimizes reprojection error for the n points.