Part II

Two-View Geometry

The Birth of Venus (detail), c. 1485 (tempera on canvas) by Sandro Botticelli (1444/5-1510)
Galleria degli Uffizi, Florence, Italy/Bridgeman Art Library

Outline

This part of the book covers the geometry of two perspective views. These views may be acquired simultaneously as in a stereo rig, or acquired sequentially, for example by a camera moving relative to the scene. These two situations are geometrically equivalent and will not be differentiated here. Each view has an associated camera matrix, $\mathtt{P}, \mathtt{P}'$, where $'$ indicates entities associated with the second view, and a 3-space point $\mathbf{X}$ is imaged as $\mathbf{x} = \mathtt{P}\mathbf{X}$ in the first view, and $\mathbf{x}' = \mathtt{P}'\mathbf{X}$ in the second. Image points $\mathbf{x}$ and $\mathbf{x}'$ *correspond* because they are the image of the same 3-space point. There are three questions that will be addressed:

(i) **Correspondence geometry.** Given an image point $\mathbf{x}$ in the first view, how does this constrain the position of the corresponding point $\mathbf{x}'$ in the second view?

(ii) **Camera geometry (motion).** Given a set of corresponding image points $\{\mathbf{x}_i \leftrightarrow \mathbf{x}'_i\}, i = 1, \ldots, n$, what are the cameras $\mathtt{P}$ and $\mathtt{P}'$ for the two views?

(iii) **Scene geometry (structure).** Given corresponding image points $\mathbf{x} \leftrightarrow \mathbf{x}'$ and cameras $\mathtt{P}, \mathtt{P}'$, what is the position of (their pre-image) $\mathbf{X}$ in 3-space?

Chapter 9 describes the *epipolar geometry* of two views, and directly answers the first question: a point in one view defines an epipolar line in the other view on which the corresponding point lies. The epipolar geometry depends only on the cameras – their relative position and their internal parameters. It does *not* depend at all on the scene structure. The epipolar geometry is represented by a 3×3 matrix called the *fundamental matrix* $\mathtt{F}$. The anatomy of the fundamental matrix is described, and its computation from camera matrices $\mathtt{P}$ and $\mathtt{P}'$ given. It is then shown that $\mathtt{P}$ and $\mathtt{P}'$ may be computed from $\mathtt{F}$ up to a projective ambiguity of 3-space.

Chapter 10 describes one of the most important results in uncalibrated multiple view geometry – a *reconstruction* of both cameras and scene structure can be computed from image point correspondences alone; no other information is required. This answers both the second and third questions simultaneously. The reconstruction obtained from point correspondences alone is up to a projective ambiguity of 3-space, and this ambiguity can be resolved by supplying well defined additional information on the cameras or scene. In this manner an affine or metric reconstruction may be computed from uncalibrated images. The following two chapters then fill in the details and numerical algorithms for computing this reconstruction.

Chapter 11 describes methods for computing $\mathtt{F}$ from a set of corresponding image points $\{\mathbf{x}_i \leftrightarrow \mathbf{x}'_i\}$, even though the structure (3D pre-image $\mathbf{X}_i$) of these points is unknown and the camera matrices are unknown. The cameras $\mathtt{P}$ and $\mathtt{P}'$ may then be determined, up to a projective ambiguity, from the computed $\mathtt{F}$.

Chapter 12 then describes the computation of scene structure by *triangulation* given the cameras and corresponding image points – the point $\mathbf{X}$ in 3-space is computed as the intersection of rays back-projected from the corresponding points $\mathbf{x}$ and $\mathbf{x}'$ via their associated cameras $\mathtt{P}, \mathtt{P}'$. Similarly, the 3D position of other geometric entities, such as lines or conics, may also be computed given their image correspondences.

Chapter 13 covers the two-view geometry of planes. It provides an alternative answer to the first question: if scene points lie on a plane, then once the geometry of this plane is computed, the image $\mathbf{x}$ of a point in one image determines the position of $\mathbf{x}'$ in the other image. The points are related by a plane projective transformation. This chapter also describes a particularly important projective transformation between views – the *infinite homography*, which is the transformation arising from the plane at infinity.

Chapter 14 describes two-view geometry in the specialized case that the two cameras $\mathtt{P}$ and $\mathtt{P}'$ are affine. This case has a number of simplifications over the general projective case, and provides a very good approximation in many practical situations.

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

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

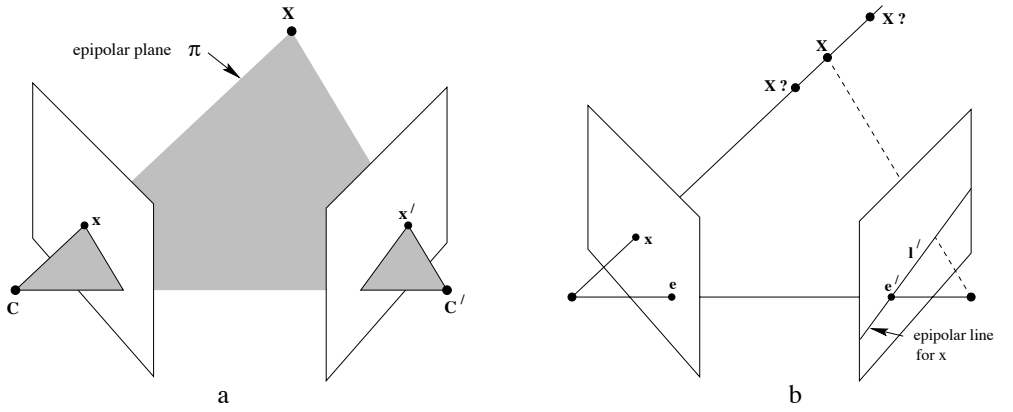


Fig. 9.1. **Point correspondence geometry.** *(a) The two cameras are indicated by their centres* $\mathbf{C}$ *and* $\mathbf{C}'$ *and image planes. The camera centres, 3-space point* $\mathbf{X}$*, and its images* $\mathbf{x}$ *and* $\mathbf{x}'$ *lie in a common plane* π*. (b) An image point* $\mathbf{x}$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

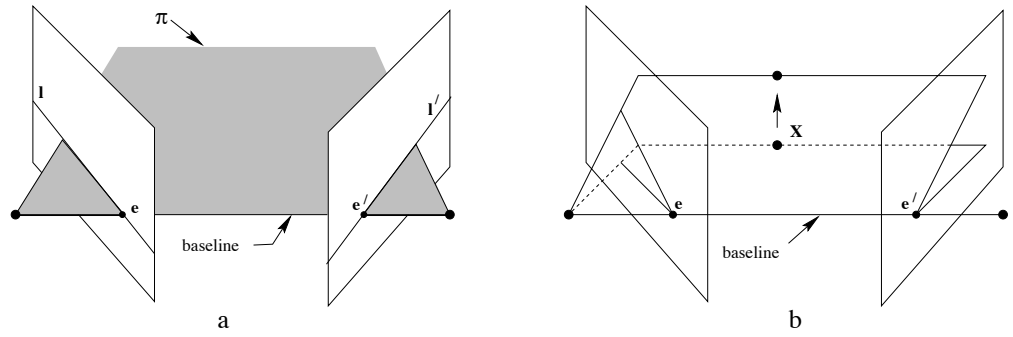


Fig. 9.2. **Epipolar geometry.** *(a) The camera baseline intersects each image plane at the epipoles* $\mathbf{e}$ *and* $\mathbf{e}'$*. Any plane* π *containing the baseline is an epipolar plane, and intersects the image planes in corresponding epipolar lines* $\mathbf{l}$ *and* $\mathbf{l}'$*. (b) As the position of the 3D point* $\mathbf{X}$ *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.*

Supposing now that we know only $\mathbf{x}$, we may ask how the corresponding point $\mathbf{x}'$ is constrained. The plane π is determined by the baseline and the ray defined by $\mathbf{x}$. From above we know that the ray corresponding to the (unknown) point $\mathbf{x}'$ lies in π, hence the point $\mathbf{x}'$ lies on the line of intersection $\mathbf{l}'$ of π 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

- 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

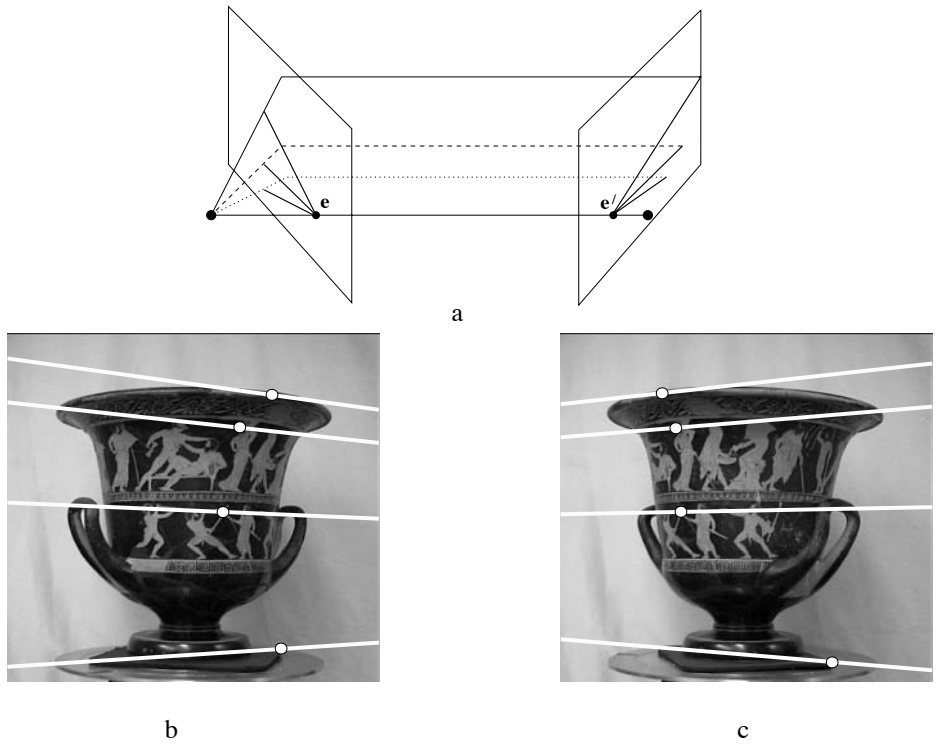


Fig. 9.3. **Converging cameras.** *(a) Epipolar geometry for converging cameras. (b) and (c) A pair of images with superimposed corresponding points and their epipolar lines (in white). The motion between the views is a translation and rotation. In each image, the direction of the other camera may be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

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).

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 $\mathbf{x}$ in one image, there exists a corresponding epipolar line $\mathbf{l}'$ in the other image. Any point $\mathbf{x}'$ in the second image matching the point $\mathbf{x}$ must lie on the epipolar line $\mathbf{l}'$. The epipolar line

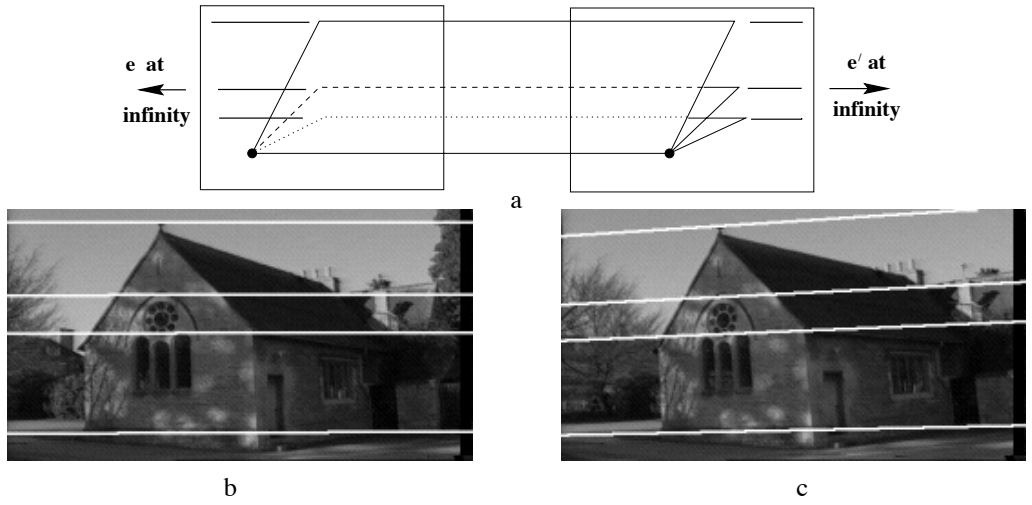


Fig. 9.4. **Motion parallel to the image plane.** *In the case of a special motion where the translation is parallel to the image plane, and the rotation axis is perpendicular to the image plane, the intersection of the baseline with the image plane is at infinity. Consequently the epipoles are at infinity, 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.*

is the projection in the second image of the ray from the point $\mathbf{x}$ through the camera centre $\mathbf{C}$ 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 $\mathtt{F}$, 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 $\mathbf{x}$ is mapped to some point $\mathbf{x}'$ in the other image lying on the epipolar line $\mathbf{l}'$. This point $\mathbf{x}'$ is a potential match for the point $\mathbf{x}$. In the second step, the epipolar line $\mathbf{l}'$ is obtained as the line joining $\mathbf{x}'$ to the epipole $\mathbf{e}'$.

Step 1: Point transfer via a plane. Refer to figure 9.5. Consider a plane $\boldsymbol{\pi}$ in space not passing through either of the two camera centres. The ray through the first camera centre corresponding to the point $\mathbf{x}$ meets the plane $\boldsymbol{\pi}$ in a point $\mathbf{X}$. This point $\mathbf{X}$ is then projected to a point $\mathbf{x}'$ in the second image. This procedure is known as transfer via the plane $\boldsymbol{\pi}$. Since $\mathbf{X}$ lies on the ray corresponding to $\mathbf{x}$, the projected point $\mathbf{x}'$ must lie on the epipolar line $\mathbf{l}'$ corresponding to the image of this ray, as illustrated in

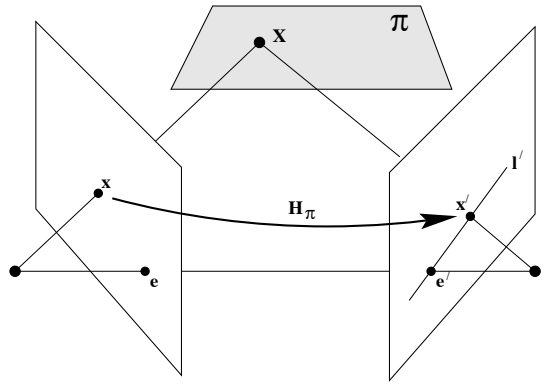


Fig. 9.5. *A point* $\mathbf{x}$ *in one image is transferred via the plane* $\boldsymbol{\pi}$ *to a matching point* $\mathbf{x}'$ *in the second image. The epipolar line through* $\mathbf{x}'$ *is obtained by joining* $\mathbf{x}'$ *to the epipole* $\mathbf{e}'$*. In symbols one may write* $\mathbf{x}'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

figure 9.1b. The points $\mathbf{x}$ and $\mathbf{x}'$ are both images of the 3D point $\mathbf{X}$ 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 $\mathtt{H}_{\boldsymbol{\pi}}$ mapping each $\mathbf{x}_i$ to $\mathbf{x}'_i$.

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

$$\mathbf{l}' = [\mathbf{e}']_\times \mathtt{H}_{\boldsymbol{\pi}}\mathbf{x} = \mathtt{F}\mathbf{x}$$

where we define $\mathtt{F} = [\mathbf{e}']_\times \mathtt{H}_{\boldsymbol{\pi}}$, the fundamental matrix. This shows

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

Geometrically, $\mathtt{F}$ 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 $\mathtt{F}$ 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, $\mathtt{P}, \mathtt{P}'$, may be derived algebraically. The following formulation is due to Xu and Zhang [Xu-96].

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

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

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

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

This is essentially the same formula for the fundamental matrix as the one derived in the previous section, the homography $\mathtt{H}_{\boldsymbol{\pi}}$ having the explicit form $\mathtt{H}_{\boldsymbol{\pi}} = \mathtt{P}'\mathtt{P}^{+}$ 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, $\mathbf{C}$ is the common camera centre of both $\mathtt{P}$ and $\mathtt{P}'$, and so $\mathtt{P}'\mathbf{C} = \mathbf{0}$. It follows that $\mathtt{F}$ 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

$$\mathtt{P} = \mathtt{K}[\mathtt{I} \mid \mathbf{0}] \qquad \mathtt{P}' = \mathtt{K}'[\mathtt{R} \mid \mathbf{t}].$$

Then

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

and

$$\begin{aligned} \mathtt{F} &= [\mathtt{P}'\mathbf{C}]_{\times}\mathtt{P}'\mathtt{P}^{+} \\ &= [\mathtt{K}'\mathbf{t}]_{\times}\mathtt{K}'\mathtt{R}\mathtt{K}^{-1} = \mathtt{K}'^{-\mathsf{T}}[\mathbf{t}]_{\times}\mathtt{R}\mathtt{K}^{-1} = \mathtt{K}'^{-\mathsf{T}}\mathtt{R}[\mathtt{R}^{\mathsf{T}}\mathbf{t}]_{\times}\mathtt{K}^{-1} = \mathtt{K}'^{-\mathsf{T}}\mathtt{R}\mathtt{K}^{\mathsf{T}}[\mathtt{K}\mathtt{R}^{\mathsf{T}}\mathbf{t}]_{\times} \end{aligned} \tag{9.2}$$

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

$$\mathbf{e} = \mathtt{P}\begin{pmatrix} -\mathtt{R}^{\mathsf{T}}\mathbf{t} \\ 1 \end{pmatrix} = \mathtt{K}\mathtt{R}^{\mathsf{T}}\mathbf{t} \qquad \mathbf{e}' = \mathtt{P}'\begin{pmatrix} \mathbf{0} \\ 1 \end{pmatrix} = \mathtt{K}'\mathbf{t}. \tag{9.3}$$

Thus we may write (9.2) as

$$\mathtt{F} = [\mathbf{e}']_{\times}\mathtt{K}'\mathtt{R}\mathtt{K}^{-1} = \mathtt{K}'^{-\mathsf{T}}[\mathbf{t}]_{\times}\mathtt{R}\mathtt{K}^{-1} = \mathtt{K}'^{-\mathsf{T}}\mathtt{R}[\mathtt{R}^{\mathsf{T}}\mathbf{t}]_{\times}\mathtt{K}^{-1} = \mathtt{K}'^{-\mathsf{T}}\mathtt{R}\mathtt{K}^{\mathsf{T}}[\mathbf{e}]_{\times}. \tag{9.4}$$

△

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 $\mathtt{F}$ 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 $\mathtt{F}$. 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*

$$\mathbf{x}'^{\mathsf{T}}\mathtt{F}\mathbf{x} = 0.$$

This is true, because if points $\mathbf{x}$ and $\mathbf{x}'$ correspond, then $\mathbf{x}'$ lies on the epipolar line $\mathbf{l}' = \mathtt{F}\mathbf{x}$ corresponding to the point $\mathbf{x}$. In other words $0 = \mathbf{x}'^{\mathsf{T}}\mathbf{l}' = \mathbf{x}'^{\mathsf{T}}\mathtt{F}\mathbf{x}$. Conversely, if image points satisfy the relation $\mathbf{x}'^{\mathsf{T}}\mathtt{F}\mathbf{x} = 0$ 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 $\mathtt{F}$ to be computed from image correspondences alone. We have seen from (9.1) that $\mathtt{F}$ may be computed from the two camera matrices, $\mathtt{P}, \mathtt{P}'$, and in particular that $\mathtt{F}$ is determined uniquely from the cameras, up to an overall scaling. However, we may now enquire how many correspondences are required to compute $\mathtt{F}$ from $\mathbf{x}'^{\mathsf{T}}\mathtt{F}\mathbf{x} = 0$, 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 $\mathtt{F}$.

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** $\mathtt{F}$ is the unique 3×3 rank 2 homogeneous matrix which satisfies

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

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.

(i) **Transpose**: If $\mathtt{F}$ is the fundamental matrix of the pair of cameras $(\mathtt{P}, \mathtt{P}')$, then $\mathtt{F}^{\mathsf{T}}$ is the fundamental matrix of the pair in the opposite order: $(\mathtt{P}', \mathtt{P})$.

(ii) **Epipolar lines**: For any point $\mathbf{x}$ in the first image, the corresponding epipolar line is $\mathbf{l}' = \mathtt{F}\mathbf{x}$. Similarly, $\mathbf{l} = \mathtt{F}^{\mathsf{T}}\mathbf{x}'$ represents the epipolar line corresponding to $\mathbf{x}'$ in the second image.

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

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

Table 9.1. *Summary of fundamental matrix properties*.

(iv) $\mathtt{F}$ 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, $\mathtt{F}$ also satisfies the constraint $\det \mathtt{F} = 0$ which removes one degree of freedom.

(v) $\mathtt{F}$ 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 $\mathbf{x}$ defines a line in the second $\mathbf{l}' = \mathtt{F}\mathbf{x}$, which is the epipolar line of $\mathbf{x}$. If $\mathbf{l}$ and $\mathbf{l}'$ are corresponding epipolar lines (see figure 9.6a) then any point $\mathbf{x}$ on $\mathbf{l}$ is mapped to the same line $\mathbf{l}'$. This means there is no inverse mapping, and $\mathtt{F}$ is not of full rank. For this reason, $\mathtt{F}$ is not a proper correlation (which would be invertible).

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 $\mathbf{e}$ in the first view, and the pencil centred at $\mathbf{e}'$ 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 $\mathbf{e}$, 2 for $\mathbf{e}'$, and 3 for the epipolar line homography which maps a line through $\mathbf{e}$ to a line through $\mathbf{e}'$. A geometric representation of this homography is given in section 9.4. Here we give an explicit formula for this mapping.

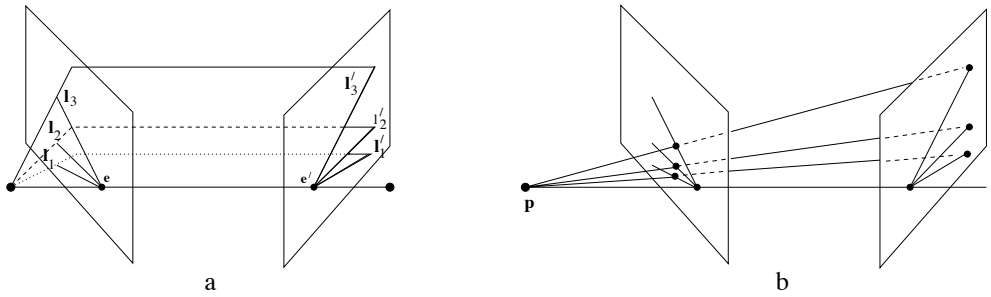


Fig. 9.6. **Epipolar line homography.** *(a) There is a pencil of epipolar lines in each image centred on the epipole. The correspondence between epipolar lines, $\mathbf{l}_i \leftrightarrow \mathbf{l}'_i$, is defined by the pencil of planes with axis the baseline. (b) The corresponding lines are related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

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

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

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

$$\mathbf{l}' = \mathtt{F}[\mathbf{e}]_\times \mathbf{l} \qquad \mathbf{l} = \mathtt{F}^\mathsf{T}[\mathbf{e}']_\times \mathbf{l}' \ .$$

9.3 Fundamental matrices arising from special motions

A special motion arises from a particular relationship between the translation direction, $\mathbf{t}$, and the direction of the rotation axis, $\mathbf{a}$. We will discuss two cases: *pure translation*, where there is no rotation; and *pure planar motion*, where $\mathbf{t}$ is orthogonal to $\mathbf{a}$ (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 $-\mathbf{t}$. In this situation points in 3-space move on straight lines parallel to $\mathbf{t}$, and the imaged intersection of these parallel lines is the vanishing point $\mathbf{v}$ in the direction of $\mathbf{t}$. This is illustrated in figure 9.7 and figure 9.8. It is evident that $\mathbf{v}$ is the epipole for both views, and the imaged parallel lines are the epipolar lines. The algebraic details are given in the following example.

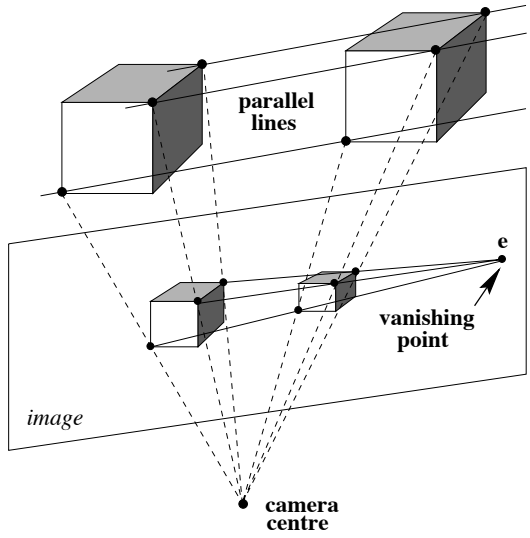


Fig. 9.7. *Under a pure translational camera motion, 3D points appear to slide along parallel rails. The images of these parallel lines intersect in a vanishing point corresponding to the translation direction. The epipole* **e** *is the vanishing point.*

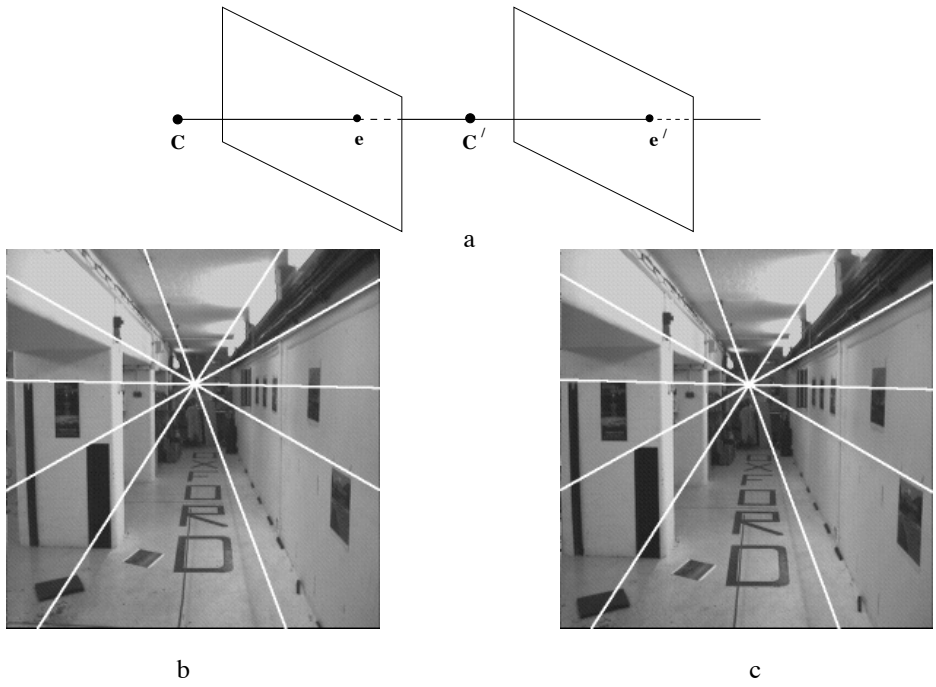


Fig. 9.8. **Pure translational motion.** *(a) under the motion the epipole is a fixed point, i.e. has the same coordinates in both images, and points appear to move along lines radiating from the epipole. The epipole in this case is termed the* Focus of Expansion *(FOE). (b) and (c) the same epipolar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

$\mathtt{P} = \mathtt{K}[\mathtt{I} \mid \mathbf{0}]$ and $\mathtt{P}' = \mathtt{K}[\mathtt{I} \mid \mathbf{t}]$. Then from (9.4) (using $\mathtt{R} = \mathtt{I}$ and $\mathtt{K} = \mathtt{K}'$)

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

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

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

The relation between corresponding points, $\mathbf{x}'^\mathsf{T}\mathtt{F}\mathbf{x} = 0$, 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 $\mathbf{x} = (x, y, 1)^\mathsf{T}$, then from $\mathbf{x} = \mathtt{P}\mathbf{X} = \mathtt{K}[\mathtt{I} \mid \mathbf{0}]\mathbf{X}$, the space point's (inhomogeneous) coordinates are $(\mathtt{X}, \mathtt{Y}, \mathtt{Z})^\mathsf{T} = \mathtt{Z}\mathtt{K}^{-1}\mathbf{x}$, where $\mathtt{Z}$ 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 $\mathbf{x}' = \mathtt{P}'\mathbf{X} = \mathtt{K}[\mathtt{I} \mid \mathbf{t}]\mathbf{X}$ that the mapping from an image point $\mathbf{x}$ to an image point $\mathbf{x}'$ is

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

The motion $\mathbf{x}' = \mathbf{x} + \mathtt{K}\mathbf{t}/\mathtt{Z}$ 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 $\mathtt{Z}$, 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 $\mathtt{F} = [\mathbf{e}']_\times$ 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 $\mathbf{l}' = \mathtt{F}\mathbf{x} = [\mathbf{e}]_\times\mathbf{x}$, and $\mathbf{x}$ lies on this line since $\mathbf{x}^\mathsf{T}[\mathbf{e}]_\times\mathbf{x} = 0$, 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 $\mathtt{H}$ 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 $\hat{\mathtt{F}} = [\mathbf{e}']_\times$, satisfying $\mathbf{x}'^\mathsf{T}\hat{\mathtt{F}}\hat{\mathbf{x}} = 0$, where $\hat{\mathbf{x}} = \mathtt{H}\mathbf{x}$ is the corrected point in the first image. From this one deduces that $\mathbf{x}'^\mathsf{T}[\mathbf{e}']_\times\mathtt{H}\mathbf{x} = 0$, and so the fundamental matrix corresponding to the initial point correspondences $\mathbf{x} \leftrightarrow \mathbf{x}'$ is $\mathtt{F} = [\mathbf{e}']_\times\mathtt{H}$. This is illustrated in figure 9.9.

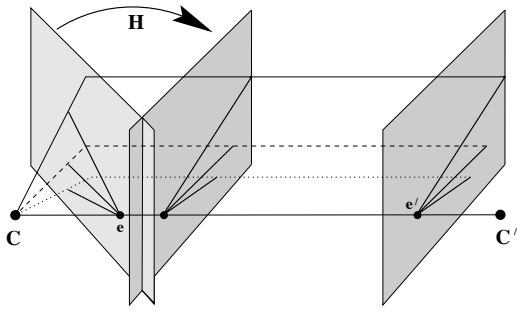


Fig. 9.9. **General camera motion.** *The first camera (on the left) may be rotated and corrected to simulate a pure translational motion. The fundamental matrix for the original pair is the product* $\mathtt{F} = [\mathbf{e}']_\times \mathtt{H}$*, where* $[\mathbf{e}']_\times$ *is the fundamental matrix of the translation, and* $\mathtt{H}$ *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 $\mathtt{P} = \mathtt{K}[\mathtt{I} \mid \mathbf{0}]$ and $\mathtt{P}' = \mathtt{K}'[\mathtt{R} \mid \mathbf{t}]$. Then as described in section 8.4.2(*p*204) the requisite projective transformation is $\mathtt{H} = \mathtt{K}'\mathtt{R}\mathtt{K}^{-1} = \mathtt{H}_\infty$, where $\mathtt{H}_\infty$ is the infinite homography (see section 13.4(*p*338)), and $\mathtt{F} = [\mathbf{e}']_\times \mathtt{H}_\infty$.

If the image point $\mathbf{x}$ is normalized as $\mathbf{x} = (x, y, 1)^\mathsf{T}$, as in example 9.6, then $(\mathtt{X}, \mathtt{Y}, \mathtt{Z})^\mathsf{T} = \mathtt{Z}\mathtt{K}^{-1}\mathbf{x}$, and from $\mathbf{x} = \mathtt{P}'\mathbf{X} = \mathtt{K}'[\mathtt{R} \mid \mathbf{t}]\mathbf{X}$ the mapping from an image point $\mathbf{x}$ to an image point $\mathbf{x}'$ is

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

The mapping is in two parts: the first term depends on the image position alone, i.e. $\mathbf{x}$, but not the point's depth $\mathtt{Z}$, 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 $\mathbf{x}$, and takes account of camera translation. In the case of pure translation ($\mathtt{R} = \mathtt{I}, \mathtt{K} = \mathtt{K}'$) (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 $\mathtt{K}' = \mathtt{K}$ then $\mathtt{F}_\mathtt{S}$, the symmetric part of $\mathtt{F}$, has rank 2 in this planar motion case (note, for a general motion the symmetric part of $\mathtt{F}$ has full rank). Thus, the condition that $\det \mathtt{F}_\mathtt{S} = 0$ is an additional constraint on $\mathtt{F}$ 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

$$\mathtt{F_S} = \left(\mathtt{F} + \mathtt{F}^\mathsf{T}\right)/2 \qquad \mathtt{F_A} = \left(\mathtt{F} - \mathtt{F}^\mathsf{T}\right)/2$$

so that $\mathtt{F} = \mathtt{F_S} + \mathtt{F_A}$.

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 $\mathbf{x}^\mathsf{T}\mathtt{F}\mathbf{x} = 0$, which is a necessary condition on corresponding points. Now, for any skew-symmetric matrix $\mathtt{A}$ the form $\mathbf{x}^\mathsf{T}\mathtt{A}\mathbf{x}$ is identically zero. Consequently only the symmetric part of $\mathtt{F}$ contributes to $\mathbf{x}^\mathsf{T}\mathtt{F}\mathbf{x} = 0$, which then reduces to $\mathbf{x}^\mathsf{T}\mathtt{F_S}\mathbf{x} = 0$. As will be seen below the matrix $\mathtt{F_S}$ 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 $\mathtt{F_S}$. We return to the horopter in chapter 22.

Symmetric part. The matrix $\mathtt{F_S}$ 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 $\mathtt{F_S}$. To see that the epipoles lie on the conic, i.e. that $\mathbf{e}^\mathsf{T}\mathtt{F_S}\mathbf{e} = 0$, start from $\mathtt{F}\mathbf{e} = \mathbf{0}$. Then $\mathbf{e}^\mathsf{T}\mathtt{F}\mathbf{e} = \mathbf{0}$ and so $\mathbf{e}^\mathsf{T}\mathtt{F_S}\mathbf{e} + \mathbf{e}^\mathsf{T}\mathtt{F_A}\mathbf{e} = 0$. However, $\mathbf{e}^\mathsf{T}\mathtt{F_A}\mathbf{e} = 0$, since for any skew-symmetric matrix $\mathtt{S}$, $\mathbf{x}^\mathsf{T}\mathtt{S}\mathbf{x} = 0$. Thus $\mathbf{e}^\mathsf{T}\mathtt{F_S}\mathbf{e} = 0$. The derivation for $\mathbf{e}'$ follows in a similar manner.

Skew-symmetric part. The matrix $\mathtt{F_A}$ is skew-symmetric and may be written as $\mathtt{F_A} = [\mathbf{x_a}]_\times$, where $\mathbf{x_a}$ is the null-vector of $\mathtt{F_A}$. 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 $\mathtt{F_S}$ is shown in figure 9.10a. The polar of $\mathbf{x_a}$ intersects the Steiner conic $\mathtt{F_S}$ 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.

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 $\mathbf{e}$ and the other through $\mathbf{e}'$. The epipolar lines are related by a 1D homography as described in section 9.2.5. The locus of intersection is the conic $\mathtt{F_S}$.

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 $\mathtt{F_S}$. 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 $\mathbf{X}_\mathrm{c}$. The point $\mathbf{X}_\mathrm{c}$ is imaged in the first view at $\mathbf{x}_\mathrm{c}$, which is the point at which $\mathbf{l}$ intersects the conic $\mathtt{F_S}$ (since $\mathtt{F_S}$ is the image of the horopter). Now the image of $\mathbf{X}_\mathrm{c}$ is also $\mathbf{x}_\mathrm{c}$ in the second view due to the fixed-point property of the horopter. So $\mathbf{x}_\mathrm{c}$ is the

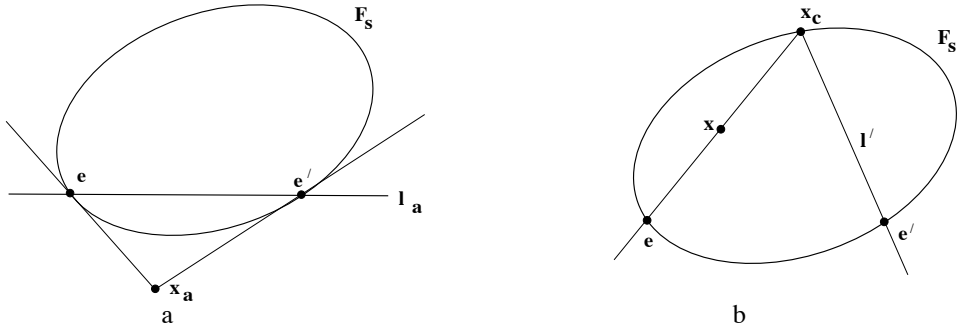


Fig. 9.10. **Geometric representation of** $\mathtt{F}$**.** *(a) The conic* $\mathtt{F_S}$ *represents the symmetric part of* $\mathtt{F}$*, and the point* $\mathtt{x_a}$ *the skew-symmetric part. The conic* $\mathtt{F_S}$ *is the locus of intersection of corresponding epipolar lines, assuming both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

image in the second view of a point on the epipolar plane of $\mathbf{x}$. It follows that $\mathbf{x}_c$ lies on the epipolar line $\mathbf{l}'$ of $\mathbf{x}$, 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 $\mathtt{F}$: 5 degrees of freedom for the conic and one each to specify the two epipoles on the conic. Given $\mathtt{F}$, then the conic $\mathtt{F_S}$, epipoles $\mathbf{e}, \mathbf{e}'$ and skew-symmetric point $\mathtt{x_a}$ are defined uniquely. However, $\mathtt{F_S}$ and $\mathtt{x_a}$ do not uniquely determine $\mathtt{F}$ since the identity of the epipoles is not recovered, i.e. the polar of $\mathtt{x_a}$ determines the epipoles but does not determine which one is $\mathbf{e}$ and which one $\mathbf{e}'$.

9.4.1 Pure planar motion

We return to the case of planar motion discussed above in section 9.3.2, where $\mathtt{F_S}$ 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:

$$\mathtt{F_S} = \mathbf{l}_h \mathbf{l}_s^\mathsf{T} + \mathbf{l}_s \mathbf{l}_h^\mathsf{T}$$

as depicted in figure 9.11a. The geometric construction of the epipolar line $\mathbf{l}'$ corresponding to a point $\mathbf{x}$ 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 $\mathbf{e}$ and $\mathbf{x}$ 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 $\mathbf{e}'$. Putting these steps together we find

$$\mathbf{l}' = \mathbf{e}' \times [\mathbf{l}_s \times (\mathbf{e} \times \mathbf{x})] = [\mathbf{e}']_\times [\mathbf{l}_s]_\times [\mathbf{e}]_\times \mathbf{x}.$$

It follows that $\mathtt{F}$ may be written as

$$\mathtt{F} = [\mathbf{e}']_\times [\mathbf{l}_s]_\times [\mathbf{e}]_\times. \tag{9.8}$$

The 6 degrees of freedom of $\mathtt{F}$ are accounted for as 2 degrees of freedom for each of the two epipoles and 2 degrees of freedom for the line.

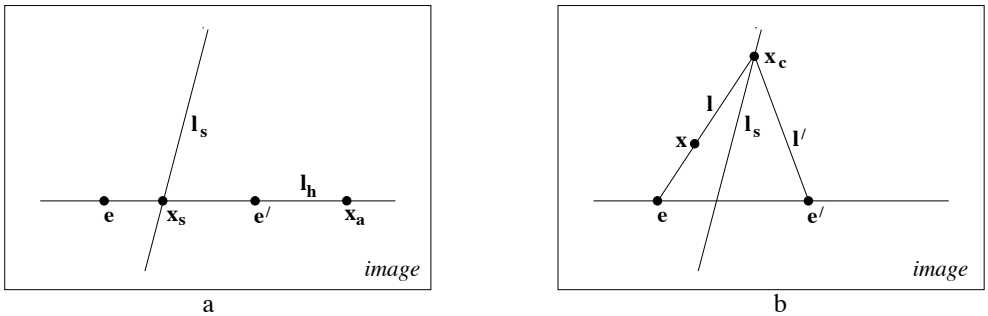


Fig. 9.11. **Geometric representation of F for planar motion.** *(a) The lines $\mathbf{l}_s$ and $\mathbf{l}_h$ constitute the Steiner conic for this motion, which is degenerate. Compare this figure with the conic for general motion shown in figure 9.10. (b) The epipolar line $\mathbf{l}'$ corresponding 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 $\mathbf{e}'$.*

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 $\mathtt{F}$ and of image relations for a point correspondence $\mathbf{x} \leftrightarrow \mathbf{x}'$. We now turn to one of the most significant properties of $\mathtt{F}$, 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 $\mathbf{l}' = \mathtt{F}\mathbf{x}$ and the correspondence condition $\mathbf{x}'^{\mathsf{T}}\mathtt{F}\mathbf{x} = 0$ 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 coordinates 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 $\hat{\mathbf{x}} = \mathtt{H}\mathbf{x}, \hat{\mathbf{x}}' = \mathtt{H}'\mathbf{x}'$, there is a corresponding map $\hat{\mathbf{l}}' = \hat{\mathtt{F}}\hat{\mathbf{x}}$ with $\hat{\mathtt{F}} = \mathtt{H}'^{-\mathsf{T}}\mathtt{F}\mathtt{H}^{-1}$ the corresponding rank 2 fundamental matrix.

Similarly, $\mathtt{F}$ only depends on projective properties of the cameras $\mathtt{P}, \mathtt{P}'$. 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. $\mathtt{F}$ does not

depend on the choice of world frame, for example a rotation of world coordinates changes $\mathtt{P}, \mathtt{P}'$, but not $\mathtt{F}$. In fact, the fundamental matrix is unchanged by a projective transformation of 3-space. More precisely,

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

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

Thus, although from (9.1–*p*244) a pair of camera matrices $(\mathtt{P}, \mathtt{P}')$ uniquely determine a fundamental matrix $\mathtt{F}$, 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 $[\mathtt{I} \mid \mathbf{0}]$, where $\mathtt{I}$ is the 3×3 identity matrix and $\mathbf{0}$ a null 3-vector. To see that this is always possible, let $\mathtt{P}$ be augmented by one row to make a 4×4 non-singular matrix, denoted $\mathtt{P}^*$. Now letting $\mathtt{H} = \mathtt{P}^{*-1}$, one verifies that $\mathtt{PH} = [\mathtt{I} \mid \mathbf{0}]$ as desired.

The following result is very frequently used

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

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

9.5.2 Projective ambiguity of cameras given $\mathtt{F}$

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

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

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

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

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

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

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

This can be tied precisely to a counting argument: the two cameras $\mathtt{P}$ and $\mathtt{P}'$ 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 $\mathtt{F}$

We have shown that $\mathtt{F}$ 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 $\mathtt{F}$. We will make use of the following characterization of the fundamental matrix $\mathtt{F}$ corresponding to a pair of camera matrices:

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

Proof. The condition that $\mathtt{P}'^\mathsf{T}\mathtt{F}\mathtt{P}$ is skew-symmetric is equivalent to $\mathbf{X}^\mathsf{T}\mathtt{P}'^\mathsf{T}\mathtt{F}\mathtt{P}\mathbf{X} = 0$ for all $\mathbf{X}$. Setting $\mathbf{x}' = \mathtt{P}'\mathbf{X}$ and $\mathbf{x} = \mathtt{P}\mathbf{X}$, this is equivalent to $\mathbf{x}'^\mathsf{T}\mathtt{F}\mathbf{x} = 0$, 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* $\mathtt{F}$ *be a fundamental matrix and* $\mathtt{S}$ *any skew-symmetric matrix. Define the pair of camera matrices*

$$\mathtt{P} = [\mathtt{I} \mid \mathbf{0}] \quad \textit{and} \quad \mathtt{P}' = [\mathtt{S}\mathtt{F} \mid \mathbf{e}'],$$

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

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

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

which is skew-symmetric.

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

As suggested by Luong and Viéville [Luong-96] a good choice for $\mathtt{S}$ is $\mathtt{S} = [\mathbf{e}']_\times$, 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* $\mathtt{F}$ *may be chosen as* $\mathtt{P} = [\mathtt{I} \mid \mathbf{0}]$ *and* $\mathtt{P}' = [[\mathbf{e}']_\times\mathtt{F} \mid \mathbf{e}']$.

Note that the camera matrix $\mathtt{P}'$ has left 3×3 submatrix $[\mathbf{e}']_\times\mathtt{F}$ 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 $\tilde{\mathtt{P}} = [\mathtt{I} \mid \mathbf{0}]$, $\tilde{\mathtt{P}}' = [\mathtt{A} + \mathbf{a}\mathbf{v}^{\mathsf{T}} \mid k\mathbf{a}]$ have the same fundamental matrix as the canonical pair, $\mathtt{P} = [\mathtt{I} \mid \mathbf{0}]$, $\mathtt{P}' = [\mathtt{A} \mid \mathbf{a}]$; 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* $\mathtt{F}$ *is given by*

$$\mathtt{P} = [\mathtt{I} \mid \mathbf{0}] \quad \mathtt{P}' = [[\mathbf{e}']_\times\mathtt{F} + \mathbf{e}'\mathbf{v}^{\mathsf{T}} \mid \lambda\mathbf{e}'] \tag{9.10}$$

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

Now, consider a pair of normalized camera matrices $\mathtt{P} = [\mathtt{I} \mid \mathbf{0}]$ and $\mathtt{P}' = [\mathtt{R} \mid \mathbf{t}]$. 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

$$\mathtt{E} = [\mathbf{t}]_\times \mathtt{R} = \mathtt{R}\,[\mathtt{R}^\mathsf{T}\mathbf{t}]_\times .$$

Definition 9.16. The defining equation for the essential matrix is

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

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

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

$$\mathtt{E} = \mathtt{K}'^\mathsf{T}\mathtt{F}\mathtt{K}. \tag{9.12}$$

9.6.1 Properties of the essential matrix

The essential matrix, $\mathtt{E} = [\mathbf{t}]_\times \mathtt{R}$, has only five degrees of freedom: both the rotation matrix $\mathtt{R}$ 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 $\mathtt{E}$ as $[\mathbf{t}]_\times \mathtt{R} = \mathtt{SR}$, where $\mathtt{S}$ is skew-symmetric. We will use the matrices

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

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

Since $\mathtt{E} = \mathtt{U}\,\mathrm{diag}(1,1,0)\mathtt{V}^\mathsf{T}$, it may seem that $\mathtt{E}$ has six degrees of freedom and not five, since both $\mathtt{U}$ and $\mathtt{V}$ 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 $\mathtt{E}$. Indeed, an alternative SVD is given by $\mathtt{E} = (\mathtt{U}\,\mathrm{diag}(\mathtt{R}_{2\times 2}, 1))\,\mathrm{diag}(1,1,0)(\mathrm{diag}(\mathtt{R}_{2\times 2}^\mathsf{T}, 1))\mathtt{V}^\mathsf{T}$ for any 2×2 rotation matrix $\mathtt{R}$.

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

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

$$\mathtt{S} = \mathtt{U}\mathtt{Z}\mathtt{U}^\mathsf{T} \quad \mathtt{R} = \mathtt{U}\mathtt{W}\mathtt{V}^\mathsf{T} \quad \textit{or} \quad \mathtt{U}\mathtt{W}^\mathsf{T}\mathtt{V}^\mathsf{T} \ . \tag{9.14}$$

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

$$\mathtt{U}\,\mathrm{diag}(1,1,0)\mathtt{V}^\mathsf{T} = \mathtt{E} = \mathtt{SR} = (\mathtt{U}\mathtt{Z}\mathtt{U}^\mathsf{T})(\mathtt{U}\mathtt{X}\mathtt{V}^\mathsf{T}) = \mathtt{U}(\mathtt{Z}\mathtt{X})\mathtt{V}^\mathsf{T}$$

from which one deduces that $\mathtt{Z}\mathtt{X} = \mathrm{diag}(1,1,0)$. Since $\mathtt{X}$ is a rotation matrix, it follows that $\mathtt{X} = \mathtt{W}$ or $\mathtt{X} = \mathtt{W}^\mathsf{T}$, as required. □

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

Result 9.19. *For a given essential matrix* $\mathtt{E} = \mathtt{U}\,\mathrm{diag}(1,1,0)\mathtt{V}^\mathsf{T}$*, and first camera matrix* $\mathtt{P} = [\mathtt{I} \mid \mathbf{0}]$*, there are four possible choices for the second camera matrix* $\mathtt{P}'$*, namely*

$$\mathtt{P}' = [\mathtt{U}\mathtt{W}\mathtt{V}^\mathsf{T} \mid +\mathbf{u}_3] \text{ or } [\mathtt{U}\mathtt{W}\mathtt{V}^\mathsf{T} \mid -\mathbf{u}_3] \text{ or } [\mathtt{U}\mathtt{W}^\mathsf{T}\mathtt{V}^\mathsf{T} \mid +\mathbf{u}_3] \text{ or } [\mathtt{U}\mathtt{W}^\mathsf{T}\mathtt{V}^\mathsf{T} \mid -\mathbf{u}_3].$$

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

$$[\mathtt{U}\mathtt{W}^\mathsf{T}\mathtt{V}^\mathsf{T} \mid \mathbf{u}_3] = [\mathtt{U}\mathtt{W}\mathtt{V}^\mathsf{T} \mid \mathbf{u}_3] \begin{bmatrix} \mathtt{V}\mathtt{W}^\mathsf{T}\mathtt{W}^\mathsf{T}\mathtt{V}^\mathsf{T} & \\ & 1 \end{bmatrix}$$

and $\mathtt{V}\mathtt{W}^\mathsf{T}\mathtt{W}^\mathsf{T}\mathtt{V}^\mathsf{T} = \mathtt{V}\,\mathrm{diag}(-1,-1,1)\mathtt{V}^\mathsf{T}$ 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 $\mathbf{X}$ 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 $\mathtt{P}'$.

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 $\mathtt{E} = [\mathbf{t}]_\times\mathtt{R}$, (i.e. including scale), and is determined only up to indeterminate scale by the equation $\mathbf{x}'^\mathsf{T}\mathtt{E}\mathbf{x} = 0$. 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 $\mathtt{E}$ 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,

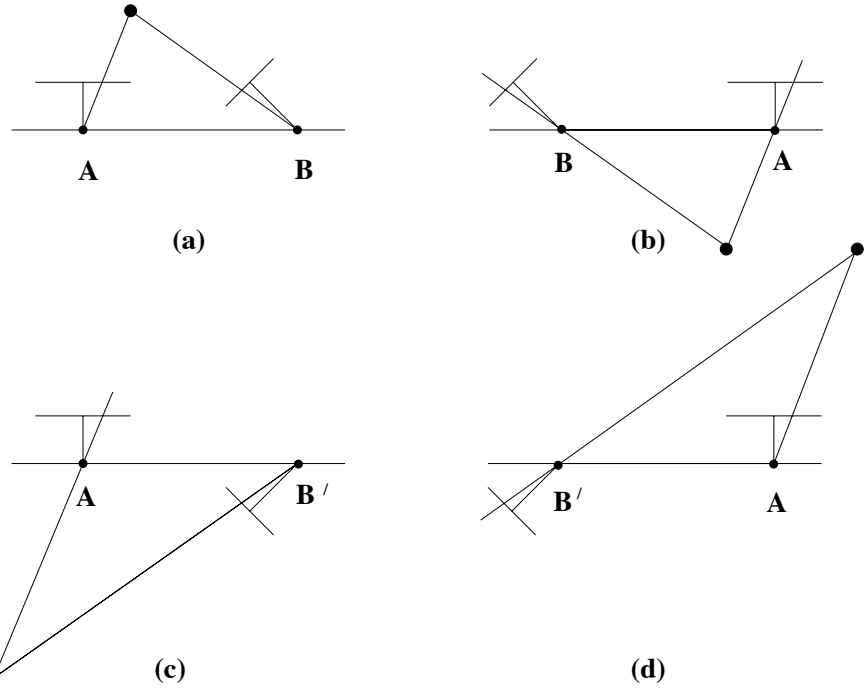


Fig. 9.12. **The four possible solutions for calibrated reconstruction from E.** *Between the left and right sides there is a baseline reversal. Between the top and bottom rows camera B rotates* 180° *about the baseline. Note, only in (a) is the reconstructed point in front of both cameras.*

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 $\mathtt{F_s}$ at the epipoles (figure 9.10a). Hint, start from $\mathtt{F}\mathbf{e} = \mathtt{F_s}\mathbf{e} + \mathtt{F_a}\mathbf{e} = \mathbf{0}$. Since $\mathbf{e}$ lies on

the conic $\mathtt{F_S}$, then $\mathbf{l}_1 = \mathtt{F_S}\mathbf{e}$ is the tangent line at $\mathbf{e}$, and $\mathbf{l}_2 = \mathtt{F_A}\mathbf{e} = [\mathbf{x_A}]_\times \mathbf{e} = \mathbf{x_A} \times \mathbf{e}$ 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 $\mathtt{K} = \mathtt{K}'$. Then from (9.12), $\mathtt{F} = \mathtt{K}^{-\mathsf{T}}\mathtt{E}\mathtt{K}^{-1}$, and so

$$\mathtt{F_S} = (\mathtt{F} + \mathtt{F}^{\mathsf{T}})/2 = \mathtt{K}^{-\mathsf{T}}(\mathtt{E} + \mathtt{E}^{\mathsf{T}})\mathtt{K}^{-1}/2 = \mathtt{K}^{-\mathsf{T}}\mathtt{E_S}\mathtt{K}^{-1}.$$

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

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

(viii) Show that the 3D points determined from one of the ambiguous reconstructions obtained from $\mathtt{E}$ 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.

10

3D Reconstruction of Cameras and Structure

This chapter describes how and to what extent the spatial layout of a scene and the cameras can be recovered from two views. Suppose that a set of image correspondences $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ are given. It is assumed that these correspondences come from a set of 3D points $\mathbf{X}_i$, which are unknown. Similarly, the position, orientation and calibration of the cameras are not known. The reconstruction task is to find the camera matrices $\mathtt{P}$ and $\mathtt{P}'$, as well as the 3D points $\mathbf{X}_i$ such that

$$\mathbf{x}_i = \mathtt{P}\mathbf{X}_i \quad \mathbf{x}'_i = \mathtt{P}'\mathbf{X}_i \quad \text{for all } i.$$

Given too few points, this task is not possible. However, if there are sufficiently many point correspondences to allow the fundamental matrix to be computed uniquely, then the scene may be reconstructed up to a projective ambiguity. This is a very significant result, and one of the major achievements of the uncalibrated approach.

The ambiguity in the reconstruction may be reduced if additional information is supplied on the cameras or scene. We describe a two-stage approach where the ambiguity is first reduced to affine, and second to metric; each stage requiring information of the appropriate class.

10.1 Outline of reconstruction method

We describe a method for reconstruction from two views as follows.

(i) Compute the fundamental matrix from point correspondences.
(ii) Compute the camera matrices from the fundamental matrix.
(iii) For each point correspondence $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$, compute the point in space that projects to these two image points.

Many variants on this method are possible. For instance, if the cameras are calibrated, then one will compute the essential matrix instead of the fundamental matrix. Furthermore, one may use information about the motion of the camera, scene constraints or partial camera calibration to obtain refinements of the reconstruction.

Each of the steps of this reconstruction method will be discussed briefly in the following paragraphs. The method described is no more than a conceptual approach to reconstruction. The reader is warned not to implement a reconstruction method based solely on the description given in this section. For real images where measurements

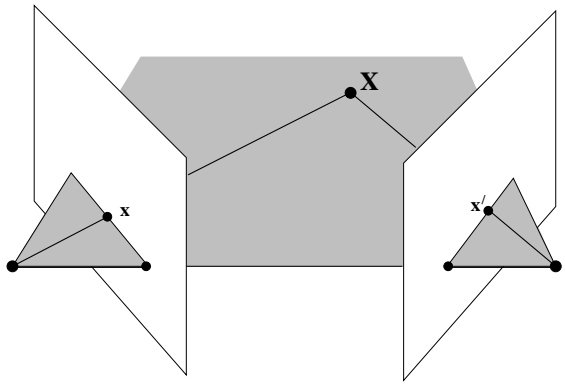


Fig. 10.1. **Triangulation.** *The image points* $\mathbf{x}$ *and* $\mathbf{x}'$ *back project to rays. If the epipolar constraint* $\mathbf{x}'^{\mathsf{T}}\mathtt{F}\mathbf{x} = 0$ *is satisfied, then these two rays lie in a plane, and so intersect in a point* $\mathbf{X}$ *in 3-space.*

are "noisy" preferred methods for reconstruction, based on this general outline, are described in chapter 11 and chapter 12.

Computation of the fundamental matrix. Given a set of correspondences $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ in two images the fundamental matrix $\mathtt{F}$ satisfies the condition $\mathbf{x}'_i\mathtt{F}\mathbf{x}_i = 0$ for all i. With the $\mathbf{x}_i$ and $\mathbf{x}'_i$ known, this equation is linear in the (unknown) entries of the matrix $\mathtt{F}$. In fact, each point correspondence generates one linear equation in the entries of $\mathtt{F}$. Given at least 8 point correspondences it is possible to solve linearly for the entries of $\mathtt{F}$ up to scale (a non-linear solution is available for 7 point correspondences). With more than 8 equations a least-squares solution is found. This is the general principle of a method for computing the fundamental matrix.

Recommended methods of computing the fundamental matrix from a set of point correspondences will be described later in chapter 11.

Computation of the camera matrices. A pair of camera matrices $\mathtt{P}$ and $\mathtt{P}'$ corresponding to the fundamental matrix $\mathtt{F}$ are easily computed using the direct formula in result 9.14.

Triangulation. Given the camera matrices $\mathtt{P}$ and $\mathtt{P}'$, let $\mathbf{x}$ and $\mathbf{x}'$ be two points in the two images that satisfy the epipolar constraint, $\mathbf{x}'^{\mathsf{T}}\mathtt{F}\mathbf{x} = 0$. As shown in chapter 9 this constraint may be interpreted geometrically in terms of the rays in space corresponding to the two image points. In particular it means that $\mathbf{x}'$ lies on the epipolar line $\mathtt{F}\mathbf{x}$. In turn this means that the two rays back-projected from image points $\mathbf{x}$ and $\mathbf{x}'$ lie in a common epipolar plane, that is, a plane passing through the two camera centres. Since the two rays lie in a plane, they will intersect in some point. This point $\mathbf{X}$ projects via the two cameras to the points $\mathbf{x}$ and $\mathbf{x}'$ in the two images. This is illustrated in figure 10.1.

The only points in 3-space that cannot be determined from their images are points on the baseline between the two cameras. In this case the back-projected rays are collinear (both being equal to the baseline) and intersect along their whole length. Thus, the point

$\mathbf{X}$ cannot be uniquely determined. Points on the baseline project to the epipoles in both images.

Numerically stable methods of actually determining the point $\mathbf{X}$ at the intersection of the two rays back-projected from $\mathbf{x}$ and $\mathbf{x}'$ will be described later in chapter 12.

10.2 Reconstruction ambiguity

In this section we discuss the inherent ambiguities involved in reconstruction of a scene from point correspondences. This topic will be discussed in a general context, without reference to a specific method of carrying out the reconstruction.

Without some knowledge of a scene's placement with respect to a 3D coordinate frame, it is generally not possible to reconstruct the absolute position or orientation of a scene from a pair of views (or in fact from any number of views). This is true independently of any knowledge which may be available about the internal parameters of the cameras, or their relative placement. For instance the exact latitude and longitude of the scene in figure 9.8(*p*248) (or any scene) cannot be computed, nor is it possible to determine whether the corridor runs north-south or east-west. This may be expressed by saying that the scene is determined at best up to a Euclidean transformation (rotation and translation) with respect to the world frame.

Only slightly less obvious is the fact that the overall scale of the scene cannot be determined. Considering figure 9.8(*p*248) once more, it is impossible based on the images alone to determine the width of the corridor. It may be two metres, one metre. It is even possible that this is an image of a doll's house and the corridor is 10 cm wide. Our common experience leads us to expect that ceilings are approximately 3m from the floor, which allows us to perceive the real scale of the scene. This extra information is an example of subsidiary knowledge of the scene not derived from image measurements. Without such knowledge therefore the scene is determined by the image only up to a similarity transformation (rotation, translation and scaling).

To give a mathematical basis to this observation, let $\mathbf{X}_i$ be a set of points and $\mathtt{P}, \mathtt{P}'$ be a pair of cameras projecting $\mathbf{X}_i$ to image points $\mathbf{x}_i$ and $\mathbf{x}'_i$. The points $\mathbf{X}_i$ and the camera pair constitute a reconstruction of the scene from the image correspondences. Now let

$$\mathtt{H}_{\mathrm{S}} = \begin{bmatrix} \mathtt{R} & \mathbf{t} \\ \mathbf{0}^{\mathsf{T}} & \lambda \end{bmatrix}$$

be any similarity transformation: $\mathtt{R}$ is a rotation, $\mathbf{t}$ a translation and λ^{-1} represents overall scaling. Replacing each point $\mathbf{X}_i$ by $\mathtt{H}_{\mathrm{S}}\mathbf{X}_i$ and cameras $\mathtt{P}$ and $\mathtt{P}'$ by $\mathtt{P}\mathtt{H}_{\mathrm{S}}^{-1}$ and $\mathtt{P}'\mathtt{H}_{\mathrm{S}}^{-1}$ respectively does not change the observed image points, since $\mathtt{P}\mathbf{X}_i = (\mathtt{P}\mathtt{H}_{\mathrm{S}}^{-1})(\mathtt{H}_{\mathrm{S}}\mathbf{X}_i)$. Furthermore, if $\mathtt{P}$ is decomposed as $\mathtt{P} = \mathtt{K}[\mathtt{R}_{\mathtt{P}} \mid \mathbf{t}_{\mathtt{P}}]$, then one computes

$$\mathtt{P}\mathtt{H}_{\mathrm{S}}^{-1} = \mathtt{K}[\mathtt{R}_{\mathtt{P}}\mathtt{R}^{-1} \mid \mathbf{t}']$$

for some $\mathbf{t}'$ that we do not need to compute more exactly. This result shows that multiplying by $\mathtt{H}_{\mathrm{S}}^{-1}$ does not change the calibration matrix of $\mathtt{P}$. Consequently this ambiguity of reconstruction exists even for calibrated cameras. It was shown by Longuet-Higgins

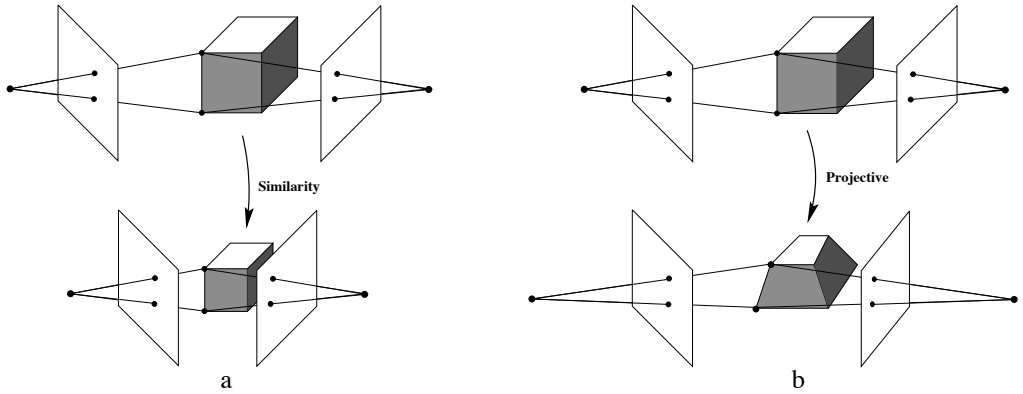


Fig. 10.2. **Reconstruction ambiguity.** *(a) If the cameras are calibrated then any reconstruction must respect the angle between rays measured in the image. A similarity transformation of the structure and camera positions does not change the measured angle. The angle between rays and the baseline (epipoles) is also unchanged. (b) If the cameras are uncalibrated then reconstructions must only respect the image points (the intersection of the rays with the image plane). A projective transformation of the structure and camera positions does not change the measured points, although the angle between rays is altered. The epipoles are also unchanged (intersection with baseline).*

([LonguetHiggins-81]) that for calibrated cameras, this is the only ambiguity of reconstruction. Thus for calibrated cameras, reconstruction is possible *up to a similarity transformation*. This is illustrated in figure 10.2a.

Projective ambiguity. If nothing is known of the calibration of either camera, nor the placement of one camera with respect to the other, then the ambiguity of reconstruction is expressed by an arbitrary projective transformation. In particular, if $\mathtt{H}$ is any 4×4 invertible matrix, representing a projective transformation of $\mathbb{P}^3$, then replacing points $\mathbf{X}_i$ by $\mathtt{H}\mathbf{X}_i$ and matrices $\mathtt{P}$ and $\mathtt{P}'$ by $\mathtt{P}\mathtt{H}^{-1}$ and $\mathtt{P}'\mathtt{H}^{-1}$ (as in the previous paragraph) does not change the image points. This shows that the points $\mathbf{X}_i$ and the cameras can be determined at best only up to a projective transformation. It is an important result, proved in this chapter (section 10.3), that this is the only ambiguity in the reconstruction of points from two images. Thus reconstruction from uncalibrated cameras is possible *up to a projective transformation*. This is illustrated in figure 10.2b.

Other types of reconstruction ambiguity result from certain assumptions on the types of motion, or partial knowledge of the cameras. For instance,

(i) If the two cameras are related via a translational motion, without change of calibration, then reconstruction is possible up to an affine transformation.
(ii) If the two cameras are calibrated apart from their focal lengths, then reconstruction is still possible up to a similarity transformation.

These two cases will be considered later in section 10.4.1 and example 19.8(*p*472), respectively.

Terminology. In any reconstruction problem derived from real data, consisting of point correspondences $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$, there exists a **true** reconstruction consisting of the actual points $\bar{\mathbf{X}}_i$ and actual cameras $\bar{\mathtt{P}}, \bar{\mathtt{P}}'$ that generated the measured observations. The

reconstructed point set $\mathbf{X}_i$ and cameras differ from the true reconstruction by a transformation belonging to a given class or group (for instance a similarity, projective or affine transformation). One speaks of *projective reconstruction*, *affine reconstruction*, *similarity reconstruction*, and so on, to indicate the type of transformation involved. However, the term *metric reconstruction* is normally used in preference to *similarity reconstruction*, being identical in meaning. The term indicates that metric properties, such as angles between lines and ratios of lengths, can be measured on the reconstruction and have their veridical values (since these are similarity invariants). In addition, the term *Euclidean reconstruction* is frequently used in the published literature to mean the same thing as a similarity or metric reconstruction, since true Euclidean reconstruction (including determination of overall scale) is not possible without extraneous information.

10.3 The projective reconstruction theorem

In this section the basic theorem of projective reconstruction from uncalibrated cameras is proved. Informally, the theorem may be stated as follows.

- *If a set of point correspondences in two views determine the fundamental matrix uniquely, then the scene and cameras may be reconstructed from these correspondences alone, and any two such reconstructions from these correspondences are projectively equivalent.*

Points lying on the line joining the two camera centres must be excluded, since such points cannot be reconstructed uniquely even if the camera matrices are determined. The formal statement is:

Theorem 10.1 (Projective reconstruction theorem). *Suppose that* $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ *is a set of correspondences between points in two images and that the fundamental matrix* $\mathtt{F}$ *is uniquely determined by the condition* $\mathbf{x}_i'^{\mathsf{T}}\mathtt{F}\mathbf{x}_i = 0$ *for all* i. *Let* $(\mathtt{P}_1, \mathtt{P}'_1, \{\mathbf{X}_{1i}\})$ *and* $(\mathtt{P}_2, \mathtt{P}'_2, \{\mathbf{X}_{2i}\})$ *be two reconstructions of the correspondences* $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$. *Then there exists a non-singular matrix* $\mathtt{H}$ *such that* $\mathtt{P}_2 = \mathtt{P}_1\mathtt{H}^{-1}$, $\mathtt{P}'_2 = \mathtt{P}'_1\mathtt{H}^{-1}$ *and* $\mathbf{X}_{2i} = \mathtt{H}\mathbf{X}_{1i}$ *for all* i, **except** *for those* i *such that* $\mathtt{F}\mathbf{x}_i = \mathbf{x}_i'^{\mathsf{T}}\mathtt{F} = \mathbf{0}$.

Proof. Since the fundamental matrix is uniquely determined by the point correspondences, one deduces that $\mathtt{F}$ is the fundamental matrix corresponding to the camera pair $(\mathtt{P}_1, \mathtt{P}'_1)$ and also to $(\mathtt{P}_2, \mathtt{P}'_2)$. According to theorem 9.10(*p*254) there is a projective transformation $\mathtt{H}$ such that $\mathtt{P}_2 = \mathtt{P}_1\mathtt{H}^{-1}$ and $\mathtt{P}'_2 = \mathtt{P}'_1\mathtt{H}^{-1}$ as required.

As for the points, one observes that $\mathtt{P}_2(\mathtt{H}\mathbf{X}_{1i}) = \mathtt{P}_1\mathtt{H}^{-1}\mathtt{H}\mathbf{X}_{1i} = \mathtt{P}_1\mathbf{X}_{1i} = \mathbf{x}_i$. On the other hand $\mathtt{P}_2\mathbf{X}_{2i} = \mathbf{x}_i$, so $\mathtt{P}_2(\mathtt{H}\mathbf{X}_{1i}) = \mathtt{P}_2\mathbf{X}_{2i}$. Thus both $\mathtt{H}\mathbf{X}_{1i}$ and $\mathbf{X}_{2i}$ map to the same point $\mathbf{x}_i$ under the action of the camera $\mathtt{P}_2$. It follows that both $\mathtt{H}\mathbf{X}_{1i}$ and $\mathbf{X}_{2i}$ lie on the same ray through the camera centre of $\mathtt{P}_2$. Similarly, it may be deduced that these two points lie on the same ray through the camera centre of $\mathtt{P}'_2$. There are two possibilities: either $\mathbf{X}_{2i} = \mathtt{H}\mathbf{X}_{1i}$ as required, or they are distinct points lying on the line joining the two camera centres. In this latter case, the image points $\mathbf{x}_i$ and $\mathbf{x}'_i$ coincide with the epipoles in the two images, and so $\mathtt{F}\mathbf{x}_i = \mathbf{x}_i'^{\mathsf{T}}\mathtt{F} = \mathbf{0}$. □

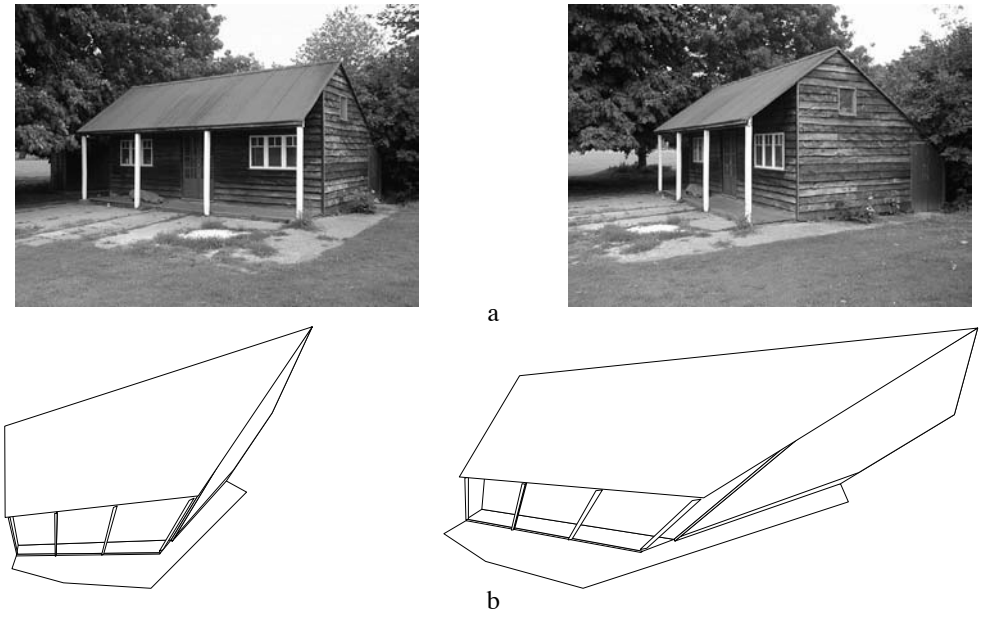


Fig. 10.3. **Projective reconstruction.** *(a) Original image pair. (b) 2 views of a 3D projective reconstruction of the scene. The reconstruction requires no information about the camera matrices, or information about the scene geometry. The fundamental matrix* $\mathtt{F}$ *is computed from point correspondences between the images, camera matrices are retrieved from* $\mathtt{F}$*, and then 3D points are computed by triangulation from the correspondences. The lines of the wireframe link the computed 3D points.*

This is an enormously significant result, since it implies that one may compute a projective reconstruction of a scene from two views based on image correspondences alone, without knowing anything about the calibration or pose of the two cameras involved. In particular the true reconstruction is within a projective transformation of the projective reconstruction. Figure 10.3 shows an example of 3D structure computed as part of a projective reconstruction from two images.

In more detail suppose the true Euclidean reconstruction is $(\mathtt{P}_{\mathrm{E}}, \mathtt{P}'_{\mathrm{E}}, \{\mathbf{X}_{\mathrm{E}i}\})$ and the projective reconstruction is $(\mathtt{P}, \mathtt{P}', \{\mathbf{X}_i\})$, then the reconstructions are related by a non-singular matrix $\mathtt{H}$ such that

$$\mathtt{P}_{\mathrm{E}} = \mathtt{P}\mathtt{H}^{-1}, \;\; \mathtt{P}'_{\mathrm{E}} = \mathtt{P}'\mathtt{H}^{-1}, \text{ and } \mathbf{X}_{\mathrm{E}i} = \mathtt{H}\mathbf{X}_i \tag{10.1}$$

where $\mathtt{H}$ is a 4×4 homography matrix which is unknown but the same for all points.

For some applications projective reconstruction is all that is required. For example, questions such as "at what point does a line intersect a plane?", "what is the mapping between two views induced by particular surfaces, such as a plane or quadric?" can be dealt with directly from the projective reconstruction. Furthermore it will be seen in the sequel that obtaining a projective reconstruction of a scene is the first step towards affine or metric reconstruction.

10.4 Stratified reconstruction

The "stratified" approach to reconstruction is to begin with a projective reconstruction and then to refine it progressively to an affine and finally a metric reconstruction, if

possible. Of course, as has just been seen, affine and metric reconstruction are not possible without further information either about the scene, the motion or the camera calibration.

10.4.1 The step to affine reconstruction

The essence of affine reconstruction is to locate the plane at infinity by some means, since this knowledge is equivalent to an affine reconstruction. This equivalence is explained in the 2D case in section 2.7(*p*47). To see this equivalence for reconstruction, suppose we have determined a projective reconstruction of a scene, consisting of a triple $(\mathtt{P}, \mathtt{P}', \{\mathbf{X}_i\})$. Suppose further that by some means a certain plane $\boldsymbol{\pi}$ has been identified as the true plane at infinity. The plane $\boldsymbol{\pi}$ is expressed as a 4-vector in the coordinate frame of the projective reconstruction. In the true reconstruction, $\boldsymbol{\pi}$ has coordinates $(0,0,0,1)^\mathsf{T}$, and we may find a projective transformation that maps $\boldsymbol{\pi}$ to $(0,0,0,1)^\mathsf{T}$. Considering the way a projective transformation acts on planes, we want to find $\mathtt{H}$ such that $\mathtt{H}^{-\mathsf{T}}\boldsymbol{\pi} = (0,0,0,1)^\mathsf{T}$. Such a transformation is given by

$$\mathtt{H} = \begin{bmatrix} \mathtt{I} & \mathbf{0} \\ \boldsymbol{\pi}^\mathsf{T} & \end{bmatrix}. \tag{10.2}$$

Indeed, it is immediately verified that $\mathtt{H}^\mathsf{T}(0,0,0,1)^\mathsf{T} = \boldsymbol{\pi}$, and thus $\mathtt{H}^{-\mathsf{T}}\boldsymbol{\pi} = (0,0,0,1)^\mathsf{T}$, as desired. The transformation $\mathtt{H}$ is now applied to all points and the two cameras. Notice, however that this formula will not work if the final coordinate of $\boldsymbol{\pi}^\mathsf{T}$ is zero. In this case, one may compute a suitable $\mathtt{H}$ by computing $\mathtt{H}^{-\mathsf{T}}$ as a Householder matrix (A4.2–*p*580) such that $\mathtt{H}^{-\mathsf{T}}\boldsymbol{\pi} = (0,0,0,1)^\mathsf{T}$.

At this point, the reconstruction that one has is not necessarily the true reconstruction – all one knows is that the plane at infinity is correctly placed. The present reconstruction differs from the true reconstruction by a projective transformation that fixes the plane at infinity. However, according to result 3.7(*p*80), a projective transformation fixing the plane at infinity is an affine transformation. Hence the reconstruction differs by an affine transformation from the true reconstruction – it is an *affine reconstruction*.

An affine reconstruction may well be sufficient for some applications. For example, the mid-point of two points and the centroid of a set of points may now be computed, and lines constructed parallel to other lines and to planes. Such computations are not possible from a projective reconstruction.

As has been stated, the plane at infinity cannot be identified unless some extra information is given. We will now give several examples of the type of information that suffices for this identification.

Translational motion

Consider the case where the camera is known to undergo a purely translational motion. In this case, it is possible to carry out affine reconstruction from two views. A simple way of seeing this is to observe that a point $\mathbf{X}$ on the plane at infinity will map to the same point in two images related by a translation. This is easily verified formally. It is also part of our common experience that as one moves in a straight line (for instance in

a car on a straight road), objects at a great distance (such as the moon) do not appear to move – only the nearby objects move past the field of view. This being so, one may invent any number of matched points $\mathbf{x}_i \leftrightarrow \mathbf{x}_i$ where a point in one image corresponds with the same point in the other image. Note that one does not actually have to observe such a correspondence in the two images – any point and the same point in the other image will do. Given a projective reconstruction, one may then reconstruct the point $\mathbf{X}_i$ corresponding to the match $\mathbf{x}_i \leftrightarrow \mathbf{x}_i$. Point $\mathbf{X}_i$ will lie on the plane at infinity. From three such points one can get three points on the plane at infinity – sufficient to determine it uniquely.

Although this argument gives a constructive proof that affine reconstruction is possible from a translating camera, this does not mean that this is the best way to proceed numerically. In fact in this case, the assumption of translational motion implies a very restricted form for the fundamental matrix – it is skew-symmetric as shown in section 9.3.1. This special form should be taken into account when solving for the fundamental matrix.

Result 10.2. *Suppose the motion of the cameras is a pure translation with no rotation and no change in the internal parameters. As shown in example 9.6(*p*249)* $\mathtt{F} = [\mathbf{e}]_\times = [\mathbf{e}']_\times$*, and for an affine reconstruction one may choose the two cameras as* $\mathtt{P} = [\mathtt{I} \mid \mathbf{0}]$ *and* $\mathtt{P}' = [\mathtt{I} \mid \mathbf{e}']$.

Scene constraints

Scene constraints or conditions may also be used to obtain an affine reconstruction. As long as three points can be identified that are known to lie on the plane at infinity, then that plane may be identified, and the reconstruction transformed to an affine reconstruction.

Parallel lines. The most obvious such condition is the knowledge that 3D lines are in reality parallel. The intersection of the two parallel lines in space gives a point on the plane at infinity. The image of this point is the vanishing point of the line, and is the point of intersection of the two imaged lines. Suppose that three sets of parallel lines can be identified in the scene. Each set intersects in a point on the plane at infinity. Provided each set has a different direction, the three points will be distinct. Since three points determine a plane, this information is sufficient to identify the plane $\boldsymbol{\pi}$.

The best way of actually computing the intersection of lines in space is a somewhat delicate problem, since in the presence of noise, lines that are intended to intersect rarely do. It is discussed in some detail in chapter 12. Correct numerical procedures for computing the plane are given in chapter 13. An example of an affine reconstruction computed from three sets of parallel scene lines is given in figure 10.4.

Note that it is not necessary to find the vanishing point in both images. Suppose the vanishing point $\mathbf{v}$ is computed from imaged parallel lines in the first image, and $\mathbf{l}'$ is a corresponding line in the second image. Vanishing points satisfy the epipolar constraint, so the corresponding vanishing point $\mathbf{v}'$ in the second image may be computed as the intersection of $\mathbf{l}'$ and the epipolar line $\mathtt{F}\mathbf{v}$ of $\mathbf{v}$. The construction of the

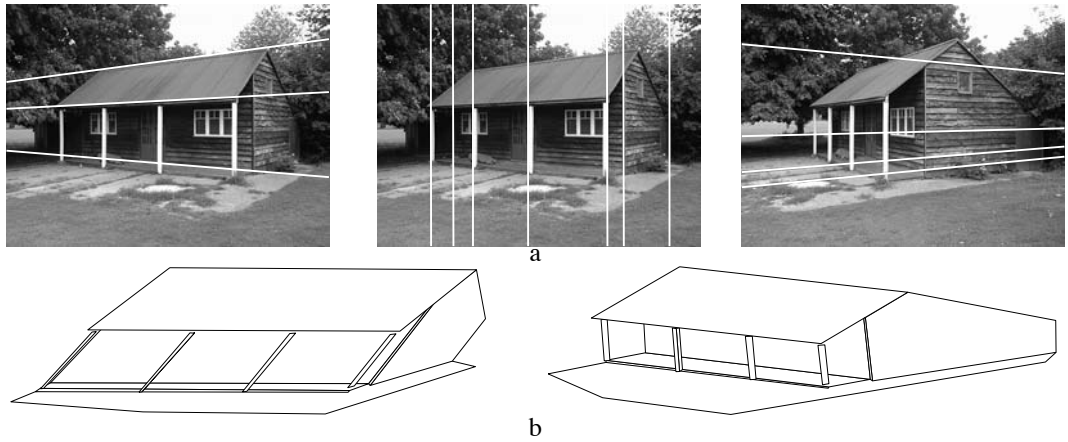


Fig. 10.4. **Affine reconstruction.** *The projective reconstruction of figure 10.3 may be upgraded to affine using parallel scene lines. (a) There are 3 sets of parallel lines in the scene, each set with a different direction. These 3 sets enable the position of the plane at infinity, $\boldsymbol{\pi}_\infty$, to be computed in the projective reconstruction. The wireframe projective reconstruction of figure 10.3 is then affinely rectified using the homography (10.2). (b) Shows two orthographic views of the wireframe affine reconstruction. Note that parallel scene lines are parallel in the reconstruction, but lines that are perpendicular in the scene are not perpendicular in the reconstruction.*

3-space point $\mathbf{X}$ can be neatly expressed algebraically as the solution of the equations $([\mathbf{v}]_\times \mathtt{P})\mathbf{X} = \mathbf{0}$ and $(\mathbf{l}'^\mathsf{T}\mathtt{P}')\mathbf{X} = 0$. These equations expresses the fact that $\mathbf{X}$ maps to $\mathbf{v}$ in the first image, and to a point on $\mathbf{l}'$ in the second image.

Distance ratios on a line. An alternative to computing vanishing points as the intersection of imaged parallel scene lines is to use knowledge of affine length ratios in the scene. For example, given two intervals on a line with a known length ratio, the point at infinity on the line may be determined. This means that from an image of a line on which a world distance ratio is known, for example that three points are equally spaced, the vanishing point may be determined. This computation, and other means of computing vanishing points and vanishing lines, are described in section 2.7(*p*47).

The infinite homography

Once the plane at infinity has been located, so that we have an affine reconstruction, then we also have an image-to-image map called the "infinite homography". This map, which is a 2D homography , is described in greater detail in chapter 13. Briefly, it is the map that transfers points from the $\mathtt{P}$ image to the $\mathtt{P}'$ image via the plane at infinity as follows: the ray corresponding to a point $\mathbf{x}$ is extended to meet the plane at infinity in a point $\mathbf{X}$; this point is projected to a point $\mathbf{x}'$ in the other image. The homography from $\mathbf{x}$ to $\mathbf{x}'$ is written as $\mathbf{x}' = \mathtt{H}_\infty\mathbf{x}$.

Having an affine reconstruction is equivalent to knowing the infinite homography as will now be shown. Given two cameras $\mathtt{P} = [\mathtt{M} \mid \mathbf{m}]$ and $\mathtt{P}' = [\mathtt{M}' \mid \mathbf{m}']$ of an affine reconstruction, the infinite homography is given by $\mathtt{H}_\infty = \mathtt{M}'\mathtt{M}^{-1}$. This is because a point $\mathbf{X} = (\tilde{\mathbf{X}}^\mathsf{T}, 0)^\mathsf{T}$ on the plane at infinity maps to $\mathbf{x} = \mathtt{M}\tilde{\mathbf{X}}$ in one image and $\mathbf{x}' = \mathtt{M}'\tilde{\mathbf{X}}$ in the other, so $\mathbf{x}' = \mathtt{M}'\mathtt{M}^{-1}\mathbf{x}$ for points on $\boldsymbol{\pi}_\infty$. Furthermore, it may be verified that

this is unchanged by a 3-space affine transformation of the cameras. Hence, the infinite homography may be computed explicitly from an affine reconstruction, and vice versa:

Result 10.3. *If an affine reconstruction has been obtained in which the camera matrices are* $\mathtt{P} = [\mathtt{I} \mid \mathbf{0}]$ *and* $\mathtt{P}' = [\mathtt{M}' \mid \mathbf{e}']$*, then the infinite homography is given by* $\mathtt{H}_\infty = \mathtt{M}'$*. Conversely, if the infinite homography* $\mathtt{H}_\infty$ *has been obtained, then the cameras of an affine reconstruction may be chosen as* $\mathtt{P} = [\mathtt{I} \mid \mathbf{0}]$ *and* $\mathtt{P}' = [\mathtt{H}_\infty \mid \mathbf{e}']$*.*

The infinite homography may be computed directly from corresponding image entities, rather than indirectly from an affine reconstruction. For example, $\mathtt{H}_\infty$ can be computed from the correspondence of three vanishing points together with $\mathtt{F}$, or the correspondence of a vanishing line and vanishing point, together with $\mathtt{F}$. The correct numerical procedure for these computations is given in chapter 13. However, such direct computations are completely equivalent to determining $\boldsymbol{\pi}_\infty$ in a projective reconstruction.

One of the cameras is affine

Another important case in which affine reconstruction is possible is when one of the cameras is known to be an affine camera as defined in section 6.3.1(*p*166). To see that this implies that affine reconstruction is possible, refer to section 6.3.5(*p*172) where it was shown that the principal plane of an affine camera is the plane at infinity. Hence to convert a projective reconstruction to an affine reconstruction, it is sufficient to find the principal plane of the camera supposed to be affine and map it to the plane $(0, 0, 0, 1)^\mathsf{T}$. Recall (section 6.2(*p*158)) that the principal plane of a camera is simply the third row of the camera matrix. For example, consider a projective reconstruction with camera matrices $\mathtt{P} = [\mathtt{I} \mid \mathbf{0}]$ and $\mathtt{P}'$, for which the first camera is supposed to be affine. To map the third row of $\mathtt{P}$ to $(0, 0, 0, 1)$ it is sufficient to swap the last two columns of the two camera matrices, while at the same time swapping the 3rd and 4th coordinates of each $\mathbf{X}_i$. This is a projective transformation corresponding to a permutation matrix $\mathtt{H}$. This shows:

Result 10.4. *Let* $(\mathtt{P}, \mathtt{P}', \{\mathbf{X}_i\})$ *be a projective reconstruction from a set of point correspondences for which* $\mathtt{P} = [\mathtt{I} \mid \mathbf{0}]$*. Suppose in truth,* $\mathtt{P}$ *is known to be an affine camera, then an affine reconstruction is obtained by swapping the last two columns of* $\mathtt{P}$ *and* $\mathtt{P}'$ *and the last two coordinates of each* $\mathbf{X}_i$*.*

Note that the condition that one of the cameras is affine places no restriction on the fundamental matrix, since any canonical camera pair $\mathtt{P} = [\mathtt{I} \mid \mathbf{0}]$ and $\mathtt{P}'$ can be transformed to a pair in which $\mathtt{P}$ is affine. If both the cameras are known to be affine, then it will be seen that the fundamental matrix has the restricted form given in (14.1–*p*345). In this case, for numerical stability, one must solve for the fundamental matrix enforcing this special form of the fundamental matrix.

Of course there is no such thing as a real affine camera – the affine camera model is an approximation which is only valid when the set of points seen in the image has small depth variation compared with the distance from the camera. Nevertheless, an assumption of an affine camera may be useful to effect the significant restriction from projective to affine reconstruction.

10.4.2 The step to metric reconstruction

Just as the key to affine reconstruction is the identification of the plane at infinity, the key to metric reconstruction is the identification of the absolute conic (section 3.6-(*p*81)). Since the absolute conic, Ω_∞, is a planar conic, lying in the plane at infinity, identifying the absolute conic implies identifying the plane at infinity.

In a stratified approach, one proceeds from projective to affine to metric reconstruction, so one knows the plane at infinity before finding the absolute conic. Suppose one has identified the absolute conic on the plane at infinity. In principle the next step is to apply an affine transformation to the affine reconstruction such that the identified absolute conic is mapped to the absolute conic in the standard Euclidean frame (it will then have the equation $X_1^2 + X_2^2 + X_3^2 = 0$, on $\boldsymbol{\pi}_\infty$). The resulting reconstruction is then related to the true reconstruction by a projective transformation which fixes the absolute conic. It follows from result 3.9(*p*82) that the projective transformation is a similarity transformation, so we have achieved a metric reconstruction.

In practice the easiest way to accomplish this is to consider the image of the absolute conic in one of the images. The image of the absolute conic (as any conic) is a conic in the image. The back-projection of this conic is a cone, which will meet the plane at infinity in a single conic, which therefore defines the absolute conic. Remember that the image of the absolute conic is a property of the image itself, and like any image point, line or other feature, is not dependent on any particular reconstruction, hence it is unchanged by 3D transformations of the reconstruction.

Suppose that in the affine reconstruction the image of the absolute conic as seen by the camera with matrix $\mathtt{P} = [\mathtt{M} \mid \mathbf{m}]$ is a conic $\boldsymbol{\omega}$. We will show how $\boldsymbol{\omega}$ may be used to define the homography $\mathtt{H}$ which transforms the affine reconstruction to a metric reconstruction:

Result 10.5. *Suppose that the image of the absolute conic is known in some image to be $\boldsymbol{\omega}$, and one has an affine reconstruction in which the corresponding camera matrix is given by $\mathtt{P} = [\mathtt{M} \mid \mathtt{m}]$. Then, the affine reconstruction may be transformed to a metric reconstruction by applying a 3D transformation of the form*

$$\mathtt{H} = \begin{bmatrix} \mathtt{A}^{-1} & \\ & 1 \end{bmatrix}$$

where $\mathtt{A}$ is obtained by Cholesky factorization from the equation $\mathtt{A}\mathtt{A}^\mathsf{T} = (\mathtt{M}^\mathsf{T}\boldsymbol{\omega}\mathtt{M})^{-1}$.

Proof. Under the transformation $\mathtt{H}$, the camera matrix $\mathtt{P}$ is transformed to a matrix $\mathtt{P}_\mathrm{M} = \mathtt{P}\mathtt{H}^{-1} = [\mathtt{M}_\mathrm{M} \mid \mathbf{m}_\mathrm{M}]$. If $\mathtt{H}^{-1}$ is of the form

$$\mathtt{H}^{-1} = \begin{bmatrix} \mathtt{A} & \mathbf{0} \\ \mathbf{0}^\mathsf{T} & 1 \end{bmatrix}$$

then $\mathtt{M}_\mathrm{M} = \mathtt{M}\mathtt{A}$. However, the image of the absolute conic is related to the camera matrix $\mathtt{P}_\mathrm{M}$ of a Euclidean frame by the relationship

$$\boldsymbol{\omega}^* = \mathtt{M}_\mathrm{M}\mathtt{M}_\mathrm{M}^\mathsf{T} \ .$$

This is because the camera matrix may be decomposed as $\mathtt{M}_{\mathrm{M}} = \mathtt{KR}$, and from (8.11–$p$210) $\boldsymbol{\omega}^* = \boldsymbol{\omega}^{-1} = \mathtt{KK}^\mathsf{T}$. Combining this with $\mathtt{M}_{\mathrm{M}} = \mathtt{MA}$ gives $\boldsymbol{\omega}^{-1} = \mathtt{MAA}^\mathsf{T}\mathtt{M}^\mathsf{T}$, which may be rearranged as $\mathtt{AA}^\mathsf{T} = (\mathtt{M}^\mathsf{T}\boldsymbol{\omega}\mathtt{M})^{-1}$. A particular value of $\mathtt{A}$ that satisfies this relationship is found by taking the Cholesky factorization of $(\mathtt{M}^\mathsf{T}\boldsymbol{\omega}\mathtt{M})^{-1}$. This latter matrix is guaranteed to be positive-definite (see result A4.5(p582)), otherwise no such matrix $\mathtt{A}$ will exist, and metric reconstruction will not be possible. □

This approach to metric reconstruction relies on identifying the image of the absolute conic. There are various ways of doing this and these are discussed next. Three sources of constraint on the image of the absolute conic are given, and in practice a combination of these constraints is used.

1. Constraints arising from scene orthogonality. Pairs of vanishing points, $\mathbf{v}_1$ and $\mathbf{v}_2$, arising from orthogonal scene lines place a single linear constraint on $\boldsymbol{\omega}$:

$$\mathbf{v}_1^\mathsf{T}\boldsymbol{\omega}\mathbf{v}_2 = 0.$$

Similarly, a vanishing point $\mathbf{v}$ and a vanishing line $\mathbf{l}$ arising from a direction and plane which are orthogonal place two constraints on $\boldsymbol{\omega}$:

$$\mathbf{l} = \boldsymbol{\omega}\mathbf{v}.$$

A common example is the vanishing point for the vertical direction and a vanishing line from the horizontal ground plane. Finally an imaged scene plane containing metric information, such as a square grid, places two constraints on $\boldsymbol{\omega}$.

2. Constraints arising from known internal parameters. If the calibration matrix of a camera is equal to $\mathtt{K}$, then the image of the absolute conic is $\boldsymbol{\omega} = \mathtt{K}^{-\mathsf{T}}\mathtt{K}^{-1}$. Thus, knowledge of the internal parameters (6.10–p157) contained in $\mathtt{K}$ may be used to constrain or determine the elements of $\boldsymbol{\omega}$. In the case where $\mathtt{K}$ is known to have zero skew ($s = 0$),

$$\omega_{12} = \omega_{21} = 0$$

and if the pixels are square (zero skew and $\alpha_x = \alpha_y$) then

$$\omega_{11} = \omega_{22}.$$

These first two sources of constraint are discussed in detail in section 8.8(p223) on single view calibration, where examples are given of calibrating a camera solely from such information. Here there is an additional source of constraints available arising from the multiple views.

3. Constraints arising from the same cameras in all images. One of the properties of the absolute conic is that its projection into an image depends only on the calibration matrix of the camera, and not on the position or orientation of the camera. In the case where both cameras $\mathtt{P}$ and $\mathtt{P}'$ have the same calibration matrix (usually meaning that both the images were taken with the same camera with different pose) one has that $\boldsymbol{\omega} = \boldsymbol{\omega}'$, that is the image of the absolute conic is the same in both images. Given

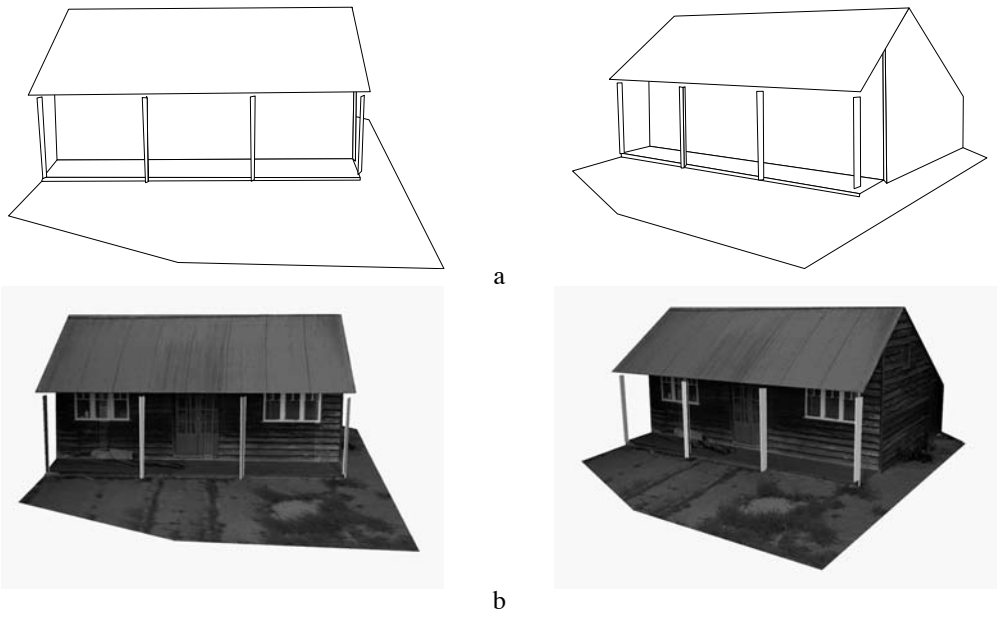


Fig. 10.5. **Metric reconstruction.** *The affine reconstruction of figure 10.4 is upgraded to metric by computing the image of the absolute conic. The information used is the orthogonality of the directions of the parallel line sets shown in figure 10.4, together with the constraint that both images have square pixels. The square pixel constraint is transferred from one image to the other using* $\mathtt{H}_\infty$. *(a) Two views of the metric reconstruction. Lines which are perpendicular in the scene are perpendicular in the reconstruction and also the aspect ratio of the sides of the house is veridical. (b) Two views of a texture mapped piecewise planar model built from the wireframes.*

sufficiently many images, one may use this property to obtain a metric reconstruction from an affine reconstruction. This method of metric reconstruction, and its use for self-calibration of a camera, will be treated in greater detail in chapter 19. For now, we give just the general principle.

Since the absolute conic lies on the plane at infinity, its image may be transferred from one view to the other via the infinite homography. This implies an equation (see result 2.13(*p*37))

$$\boldsymbol{\omega}' = \mathtt{H}_\infty^{-\mathsf{T}} \boldsymbol{\omega} \mathtt{H}_\infty^{-1} \tag{10.3}$$

where $\boldsymbol{\omega}$ and $\boldsymbol{\omega}'$ are images of Ω_∞ in the two views. In forming these equations it is necessary to have an affine reconstruction already, since the infinite homography must be known. If $\boldsymbol{\omega} = \boldsymbol{\omega}'$, then (10.3) gives a set of linear equations in the entries of $\boldsymbol{\omega}$. In general this set of linear equations places four constraints on $\boldsymbol{\omega}$, and since $\boldsymbol{\omega}$ has 5 degrees of freedom it is not completely determined. However, by combining these linear equations with those above provided by scene orthogonality or known internal parameters, $\boldsymbol{\omega}$ may be determined uniquely. Indeed (10.3) may be used to transfer constraints on $\boldsymbol{\omega}$ to constraints on $\boldsymbol{\omega}'$. Figure 10.5 shows an example of a metric reconstruction computed by combining constraints in this manner.

10.4.3 Direct metric reconstruction using $\boldsymbol{\omega}$

The previous discussion showed how knowledge of the image of the absolute conic (IAC) may be used to transform an affine to a metric reconstruction. However, knowing $\boldsymbol{\omega}$ it is possible to proceed directly to metric reconstruction, given at least two views. This can be accomplished in at least two different ways. The most evident approach is to use the IAC to compute calibration of each of the cameras, and then carry out a calibrated reconstruction.

This method relies on the connection of $\boldsymbol{\omega}$ to the calibration matrix $\mathtt{K}$, namely $\boldsymbol{\omega} = (\mathtt{K}\mathtt{K}^{\mathsf{T}})^{-1}$. Thus one can compute $\mathtt{K}$ from $\boldsymbol{\omega}$ by inverting it and then applying Cholesky factorization to obtain $\mathtt{K}$. If the IAC is known in each image, then both cameras may be calibrated in this way. Next with calibrated cameras, a metric reconstruction of the scene may be computed using the essential matrix, as in section 9.6. Note that four possible solutions may result. Two of these are just mirror images, but the other two are different, forming a twisted pair. (Though all solutions but one may be ruled out by consideration of points lying in front of the cameras.)

A more conceptual approach to metric reconstruction is to use knowledge of the IAC to directly determine the plane at infinity and the absolute conic. Knowing the camera matrices $\mathtt{P}$ and $\mathtt{P}'$ in a projective frame, and a conic (specifically the image of the absolute conic) in each image, then Ω_∞ may be explicitly computed in 3-space. This is achieved by back-projecting the conics to cones, which must intersect in the absolute conic. Thus, Ω_∞ and its support plane $\boldsymbol{\pi}_\infty$ are determined (see exercise (x) on page 342 for an algebraic solution). However, two cones will in general intersect in two different plane conics, each lying in a different support plane. Thus there are two possible solutions for the absolute conic, which one can identify as belonging to the two different reconstructions constituting the twisted pair ambiguity.

10.5 Direct reconstruction – using ground truth

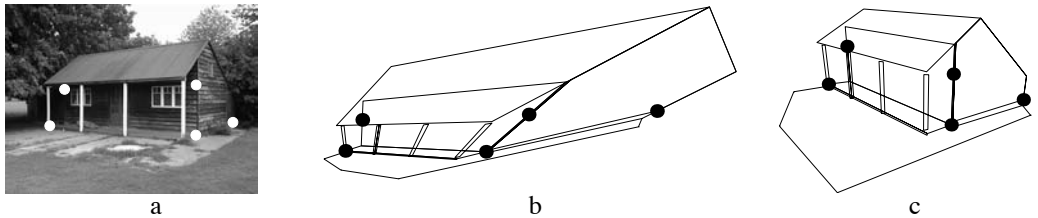
It is possible to jump directly from a projective reconstruction to a metric reconstruction if "ground control points" (that is points with known 3D locations in a Euclidean world frame) are given. Suppose we have a set of n such ground control points $\{\mathbf{X}_{\mathrm{E}i}\}$ which are imaged at $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$. We wish to use these points to transform the projective reconstruction to metric.

The 3D location $\{\mathbf{X}_i\}$ of the control points in the projective reconstruction may be computed from their image correspondences $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$. Since the projective reconstruction is related by a homography to the true reconstruction we then have from (10.1) the equations:

$$\mathbf{X}_{\mathrm{E}i} = \mathtt{H}\mathbf{X}_i, \quad i = 1, \ldots, n.$$

Each point correspondence provides 3 linearly independent equations on the elements of $\mathtt{H}$, and since $\mathtt{H}$ has 15 degrees of freedom a linear solution is obtained provided $n \geq 5$ (and no four of the control points are coplanar). This computation, and the proper numerical procedures, are described in chapter 4.

Alternatively, one may bypass the computation of the $\mathbf{X}_i$ and compute $\mathtt{H}$ by relating



a b c

Fig. 10.6. **Direct reconstruction.** *The projective reconstruction of figure 10.3 may be upgraded to metric by specifying the position of five (or more) world points: (a) the five points used; (b) the corresponding points on the projective reconstruction of figure 10.3; (c) the reconstruction after the five points are mapped to their world positions.*

the known ground control points directly to image measurements. Thus as in the DLT algorithm for camera resection (section 7.1(p178)) the equation

$$\mathbf{x}_i = \mathtt{P}\mathtt{H}^{-1}\mathbf{X}_{\mathrm{E}i}$$

provides two linearly independent equations in the entries of the unknown $\mathtt{H}^{-1}$, all other quantities being known. Similarly equations may be derived from the other image if $\mathbf{x}'_i$ is known. It is not necessary for the ground control points to be visible in both images. Note however that if both $\mathbf{x}_i$ and $\mathbf{x}'_i$ *are* visible for a given control point $\mathbf{X}_{\mathrm{E}i}$ then because of the coplanarity constraint on $\mathbf{x}$ and $\mathbf{x}'$, the four equations generated in this way contain only three independent ones.

Once $\mathtt{H}$ has been computed it may be used to transform the cameras $\mathtt{P}, \mathtt{P}'$ of the projective reconstruction to their true Euclidean counterparts. An example of metric structure computed by this direct method is shown in figure 10.6.

10.6 Closure

In this chapter we have overviewed the steps necessary to produce a metric reconstruction from a pair of images. This overview is summarized in algorithm 10.1, and the computational procedures for these steps are described in the following chapters. As usual the general discussion has been restricted mainly to points, but the ideas (triangulation, ambiguity, stratification) apply equally to other image features such as lines, conics etc.

It has been seen that for a metric reconstruction it is necessary to identify two entities in the projective frame; these are the plane at infinity $\boldsymbol{\pi}_\infty$ (for affine), together with the absolute conic Ω_∞ (for metric). Conversely, given $\mathtt{F}$ and a pair of calibrated cameras then $\boldsymbol{\pi}_\infty$ and Ω_∞ may be explicitly computed in 3-space. These entities each have an image-based counterpart: specification of the infinite homography, $\mathtt{H}_\infty$, is equivalent to specifying $\boldsymbol{\pi}_\infty$ in 3-space; and specifying the image of the absolute conic, $\boldsymbol{\omega}$, in each view is equivalent to specifying $\boldsymbol{\pi}_\infty$ and Ω_∞ in 3-space. This equivalence is summarized in table 10.1.

Finally, it is worth noting that if metric precision is not the goal then an acceptable metric reconstruction is generally obtained directly from the projective if approximately correct internal parameters are guessed. Such a "quasi-Euclidean reconstruction" is often suitable for visualization purposes.

Objective

Given two uncalibrated images compute a metric reconstruction $(\mathtt{P}_{\mathrm{M}}, \mathtt{P}'_{\mathrm{M}}, \{\mathbf{X}_{\mathrm{M}i}\})$ of the cameras and scene structure, i.e. a reconstruction that is within a similarity transformation of the true cameras and scene structure.

Algorithm

(i) **Compute a projective reconstruction** $(\mathtt{P}, \mathtt{P}', \{\mathbf{X}_i\})$**:**

(a) **Compute the fundamental matrix** from point correspondences $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ between the images.

(b) **Camera retrieval:** compute the camera matrices $\mathtt{P}, \mathtt{P}'$ from the fundamental matrix.

(c) **Triangulation:** for each point correspondence $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$, compute the point $\mathbf{X}_i$ in space that projects to these two image points.

(ii) **Rectify the projective reconstruction to metric:**

- either **Direct method:** Compute the homography $\mathtt{H}$ such that $\mathbf{X}_{\mathrm{E}i} = \mathtt{H}\mathbf{X}_i$ from five or more ground control points $\mathbf{X}_{\mathrm{E}i}$ with known Euclidean positions. Then the metric reconstruction is

$$\mathtt{P}_{\mathrm{M}} = \mathtt{P}\mathtt{H}^{-1}, \;\; \mathtt{P}'_{\mathrm{M}} = \mathtt{P}'\mathtt{H}^{-1}, \;\; \mathbf{X}_{\mathrm{M}i} = \mathtt{H}\mathbf{X}_i.$$

- or **Stratified method:**

(a) **Affine reconstruction:** Compute the plane at infinity, $\boldsymbol{\pi}_\infty$, as described in section 10.4.1, and then upgrade the projective reconstruction to an affine reconstruction with the homography

$$\mathtt{H} = \begin{bmatrix} \mathtt{I} & \mathbf{0} \\ \boldsymbol{\pi}_\infty^{\mathsf{T}} & \end{bmatrix}.$$

(b) **Metric reconstruction:** Compute the image of the absolute conic, $\boldsymbol{\omega}$, as described in section 10.4.2, and then upgrade the affine reconstruction to a metric reconstruction with the homography

$$\mathtt{H} = \begin{bmatrix} \mathtt{A}^{-1} & \\ & 1 \end{bmatrix}$$

where $\mathtt{A}$ is obtained by Cholesky factorization from the equation $\mathtt{A}\mathtt{A}^{\mathsf{T}} = (\mathtt{M}^{\mathsf{T}}\boldsymbol{\omega}\mathtt{M})^{-1}$, and $\mathtt{M}$ is the first 3×3 submatrix of the camera in the affine reconstruction for which $\boldsymbol{\omega}$ is computed.

Algorithm 10.1. *Computation of a metric reconstruction from two uncalibrated images.*

10.6.1 The literature

Koenderink and van Doorn [Koenderink-91] give a very elegant discussion of stratification for affine cameras. This was extended to perspective in [Faugeras-95b], with developments given by Luong and Viéville [Luong-94, Luong-96]. The possibility of projective reconstruction given $\mathtt{F}$ appeared in [Faugeras-92b] and Hartley *et al*. [Hartley-92c].

The method of computing affine reconstruction from pure translation first appeared in Moons *et al*. [Moons-94]. Combining scene and internal parameter constraints over multiple views is described in [Faugeras-95c, Liebowitz-99b, Sturm-99c].

Image information provided	View relations and projective objects	3-space objects	Reconstruction ambiguity
Point correspondences	$\mathtt{F}$		Projective
Point correspondences including vanishing points	$\mathtt{F}, \mathtt{H}_\infty$	$\boldsymbol{\pi}_\infty$	Affine
Point correspondences and internal camera calibration	$\mathtt{F}, \mathtt{H}_\infty$ $\boldsymbol{\omega}, \boldsymbol{\omega}'$	$\boldsymbol{\pi}_\infty$ Ω_∞	Metric

Table 10.1. *The two-view relations, image entities, and their 3-space counterparts for various classes of reconstruction ambiguity.*

10.6.2 Notes and exercises

(i) Using only (implicit) image relations (i.e. without an explicit 3D reconstruction) and given the images of a line $\mathbf{L}$ and point $\mathbf{X}$ (not on $\mathbf{L}$) in two views, together with $\mathtt{H}_\infty$ between the views, compute the image of the line in 3-space parallel to $\mathbf{L}$ and through $\mathbf{X}$. Other examples of this implicit approach to computation are given in [Zeller-96].

11

Computation of the Fundamental Matrix F

This chapter describes numerical methods for estimating the fundamental matrix given a set of point correspondences between two images. We begin by describing the equations on F generated by point correspondences in two images, and their minimal solution. The following sections then give linear methods for estimating F using algebraic distance, and then various geometric cost functions and solution methods including the MLE ("Gold Standard") algorithm, and Sampson distance.

An algorithm is then described for automatically obtaining point correspondences, so that F may be estimated directly from an image pair. We discuss the estimation of F for special camera motions.

The chapter also covers a method of image rectification based on the computed F.

11.1 Basic equations

The fundamental matrix is defined by the equation

$$\mathbf{x}'^{\mathsf{T}}\mathtt{F}\mathbf{x} = 0 \tag{11.1}$$

for any pair of matching points $\mathbf{x} \leftrightarrow \mathbf{x}'$ in two images. Given sufficiently many point matches $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ (at least 7), equation (11.1) can be used to compute the unknown matrix F. In particular, writing $\mathbf{x} = (x, y, 1)^{\mathsf{T}}$ and $\mathbf{x}' = (x', y', 1)^{\mathsf{T}}$ each point match gives rise to one linear equation in the unknown entries of F. The coefficients of this equation are easily written in terms of the known coordinates $\mathbf{x}$ and $\mathbf{x}'$. Specifically, the equation corresponding to a pair of points $(x, y, 1)$ and $(x', y', 1)$ is

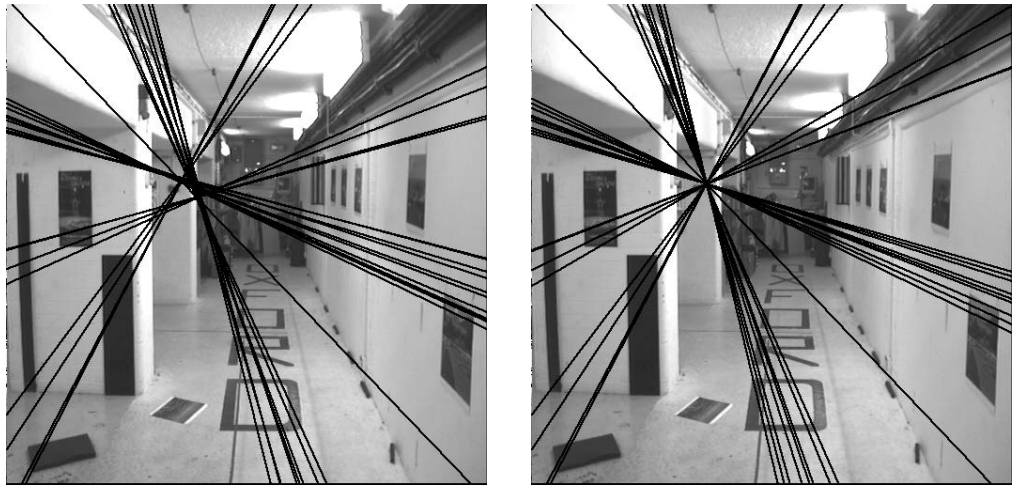
$$x'xf_{11} + x'yf_{12} + x'f_{13} + y'xf_{21} + y'yf_{22} + y'f_{23} + xf_{31} + yf_{32} + f_{33} = 0. \tag{11.2}$$

Denote by $\mathbf{f}$ the 9-vector made up of the entries of F in row-major order. Then (11.2) can be expressed as a vector inner product

$$(x'x, x'y, x', y'x, y'y, y', x, y, 1)\,\mathbf{f} = 0.$$

From a set of n point matches, we obtain a set of linear equations of the form

$$\mathtt{A}\mathbf{f} = \begin{bmatrix} x'_1x_1 & x'_1y_1 & x'_1 & y'_1x_1 & y'_1y_1 & y'_1 & x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ x'_nx_n & x'_ny_n & x'_n & y'_nx_n & y'_ny_n & y'_n & x_n & y_n & 1 \end{bmatrix} \mathbf{f} = \mathbf{0}. \tag{11.3}$$



a b

Fig. 11.1. **Epipolar lines.** *(a) the effect of a non-singular fundamental matrix. Epipolar lines computed as* $\mathbf{l}' = \mathtt{F}\mathbf{x}$ *for varying* $\mathbf{x}$ *do not meet in a common epipole. (b) the effect of enforcing singularity using the SVD method described here.*

This is a homogeneous set of equations, and $\mathbf{f}$ can only be determined up to scale. For a solution to exist, matrix $\mathtt{A}$ must have rank at most 8, and if the rank is exactly 8, then the solution is unique (up to scale), and can be found by linear methods – the solution is the generator of the right null-space of $\mathtt{A}$.

If the data is not exact, because of noise in the point coordinates, then the rank of $\mathtt{A}$ may be greater than 8 (in fact equal to 9, since $\mathtt{A}$ has 9 columns). In this case, one finds a least-squares solution. Apart from the specific form of the equations (compare (11.3) with (4.3–*p*89)) the problem is essentially the same as the estimation problem considered in section 4.1.1(*p*90). Refer to the algorithm 4.1(*p*91). The least-squares solution for $\mathbf{f}$ is the singular vector corresponding to the smallest singular value of $\mathtt{A}$, that is, the last column of $\mathtt{V}$ in the SVD $\mathtt{A} = \mathtt{U}\mathtt{D}\mathtt{V}^{\mathsf{T}}$. The solution vector $\mathbf{f}$ found in this way minimizes $\|\mathtt{A}\mathbf{f}\|$ subject to the condition $\|\mathbf{f}\| = 1$. The algorithm just described is the essence of a method called the 8-point algorithm for computation of the fundamental matrix.

11.1.1 The singularity constraint

An important property of the fundamental matrix is that it is singular, in fact of rank 2. Furthermore, the left and right null-spaces of $\mathtt{F}$ are generated by the vectors representing (in homogeneous coordinates) the two epipoles in the two images. Most applications of the fundamental matrix rely on the fact that it has rank 2. For instance, if the fundamental matrix is not singular then computed epipolar lines are not coincident, as is demonstrated by figure 11.1. The matrix $\mathtt{F}$ found by solving the set of linear equations (11.3) will not in general have rank 2, and we should take steps to enforce this constraint. The most convenient way to do this is to correct the matrix $\mathtt{F}$ found by the SVD solution from $\mathtt{A}$. Matrix $\mathtt{F}$ is replaced by the matrix $\mathtt{F}'$ that minimizes the Frobenius norm $\|\mathtt{F} - \mathtt{F}'\|$ subject to the condition $\det \mathtt{F}' = 0$. A convenient method of

doing this is to again use the SVD. In particular, let $\mathtt{F} = \mathtt{U}\mathtt{D}\mathtt{V}^\mathsf{T}$ be the SVD of $\mathtt{F}$, where $\mathtt{D}$ is a diagonal matrix $\mathtt{D} = \mathrm{diag}(r, s, t)$ satisfying $r \geq s \geq t$. Then $\mathtt{F}' = \mathtt{U}\mathrm{diag}(r, s, 0)\mathtt{V}^\mathsf{T}$ minimizes the Frobenius norm of $\mathtt{F} - \mathtt{F}'$.

Thus, the 8-point algorithm for computation of the fundamental matrix may be formulated as consisting of two steps, as follows.

(i) **Linear solution.** A solution $\mathtt{F}$ is obtained from the vector $\mathbf{f}$ corresponding to the smallest singular value of $\mathtt{A}$, where $\mathtt{A}$ is defined in (11.3).

(ii) **Constraint enforcement.** Replace $\mathtt{F}$ by $\mathtt{F}'$, the closest singular matrix to $\mathtt{F}$ under a Frobenius norm. This correction is done using the SVD.

The algorithm thus stated is extremely simple, and readily implemented, assuming that appropriate linear algebra routines are available. As usual normalization is required, and we return to this in section 11.2.

11.1.2 The minimum case – seven point correspondences

The equation $\mathbf{x}_i'^\mathsf{T}\mathtt{F}\mathbf{x}_i = 0$ gives rise to a set of equations of the form $\mathtt{A}\mathbf{f} = \mathbf{0}$. If $\mathtt{A}$ has rank 8, then it is possible to solve for $\mathbf{f}$ up to scale. In the case where the matrix $\mathtt{A}$ has rank seven, it is still possible to solve for the fundamental matrix by making use of the singularity constraint. The most important case is when only 7 point correspondences are known (other cases are discussed in section 11.9). This leads to a 7×9 matrix $\mathtt{A}$, which generally will have rank 7.

The solution to the equations $\mathtt{A}\mathbf{f} = \mathbf{0}$ in this case is a 2-dimensional space of the form $\alpha\mathtt{F}_1 + (1-\alpha)\mathtt{F}_2$, where α is a scalar variable. The matrices $\mathtt{F}_1$ and $\mathtt{F}_2$ are obtained as the matrices corresponding to the generators $\mathbf{f}_1$ and $\mathbf{f}_2$ of the right null-space of $\mathtt{A}$. Now, we use the constraint that $\det \mathtt{F} = 0$. This may be written as $\det(\alpha\mathtt{F}_1 + (1-\alpha)\mathtt{F}_2) = 0$. Since $\mathtt{F}_1$ and $\mathtt{F}_2$ are known, this leads to a cubic polynomial equation in α. This polynomial equation may be solved to find the value of α. There will be either one or three real solutions (the complex solutions are discarded [Hartley-94c]). Substituting back in the equation $\mathtt{F} = \alpha\mathtt{F}_1 + (1-\alpha)\mathtt{F}_2$ gives one or three possible solutions for the fundamental matrix.

This method of computing one or three fundamental matrices for the minimum number of points (seven) is used in the robust algorithm of section 11.6. We return to the issue of the number of solutions in section 11.9.

11.2 The normalized 8-point algorithm

The 8-point algorithm is the simplest method of computing the fundamental matrix, involving no more than the construction and (least-squares) solution of a set of linear equations. If care is taken, then it can perform extremely well. The original algorithm is due to Longuet-Higgins [LonguetHiggins-81]. The key to success with the 8-point algorithm is proper careful normalization of the input data before constructing the equations to solve. The subject of normalization of input data has applications to many of the algorithms of this book, and is treated in general terms in section 4.4-(p104). In the case of the 8-point algorithm, a simple transformation (translation and

Objective

Given $n \geq 8$ image point correspondences $\{\mathbf{x}_i \leftrightarrow \mathbf{x}'_i\}$, determine the fundamental matrix $\mathtt{F}$ such that $\mathbf{x}'^{\mathsf{T}}_i \mathtt{F} \mathbf{x}_i = 0$.

Algorithm

(i) **Normalization:** Transform the image coordinates according to $\hat{\mathbf{x}}_i = \mathtt{T}\mathbf{x}_i$ and $\hat{\mathbf{x}}'_i = \mathtt{T}'\mathbf{x}'_i$, where $\mathtt{T}$ and $\mathtt{T}'$ are normalizing transformations consisting of a translation and scaling.

(ii) Find the fundamental matrix $\hat{\mathtt{F}}'$ corresponding to the matches $\hat{\mathbf{x}}_i \leftrightarrow \hat{\mathbf{x}}'_i$ by

  (a) **Linear solution:** Determine $\hat{\mathtt{F}}$ from the singular vector corresponding to the smallest singular value of $\hat{\mathtt{A}}$, where $\hat{\mathtt{A}}$ is composed from the matches $\hat{\mathbf{x}}_i \leftrightarrow \hat{\mathbf{x}}'_i$ as defined in (11.3).

  (b) **Constraint enforcement:** Replace $\hat{\mathtt{F}}$ by $\hat{\mathtt{F}}'$ such that $\det \hat{\mathtt{F}}' = 0$ using the SVD (see section 11.1.1).

(iii) **Denormalization:** Set $\mathtt{F} = \mathtt{T}'^{\mathsf{T}} \hat{\mathtt{F}}' \mathtt{T}$. Matrix $\mathtt{F}$ is the fundamental matrix corresponding to the original data $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$.

Algorithm 11.1. *The normalized 8-point algorithm for* $\mathtt{F}$.

scaling) of the points in the image before formulating the linear equations leads to an enormous improvement in the conditioning of the problem and hence in the stability of the result. The added complexity of the algorithm necessary to do this transformation is insignificant.

The suggested normalization is a translation and scaling of each image so that the centroid of the reference points is at the origin of the coordinates and the RMS distance of the points from the origin is equal to $\sqrt{2}$. This is carried out for essentially the same reasons as in chapter 4. The basic method is analogous to algorithm 4.2(p109) and is summarized in algorithm 11.1.

Note that it is recommended that the singularity condition should be enforced before denormalization. For a justification of this, refer to [Hartley-97c].

11.3 The algebraic minimization algorithm

The normalized 8-point algorithm includes a method for enforcing the singularity constraint on the fundamental matrix. The initial estimate $\mathtt{F}$ is replaced by the singular matrix $\mathtt{F}'$ that minimizes the difference $\|\mathtt{F}' - \mathtt{F}\|$. This is done using the SVD, and has the advantage of being simple and rapid.

Numerically, however, this method is not optimal, since all the entries of $\mathtt{F}$ do not have equal importance, and indeed some entries are more tightly constrained by the point-correspondence data than others. A more correct procedure would be to compute a covariance matrix from the entries of $\mathtt{F}$ in terms of the input data, and then to find the singular matrix $\mathtt{F}'$ closest to $\mathtt{F}$ in terms of Mahalanobis distance with respect to this covariance. Unfortunately, minimization of the Mahalanobis distance $\|\mathtt{F} - \mathtt{F}'\|_\Sigma$ cannot be done linearly for a general covariance matrix Σ, so this approach is unattractive.

An alternative procedure is to find the desired singular matrix $\mathtt{F}'$ directly. Thus, just as $\mathtt{F}$ is computed by minimizing the norm $\|\mathtt{A}\mathbf{f}\|$ subject to $\|\mathbf{f}\| = 1$, so one should aim

to find the *singular* matrix $\mathtt{F}'$ that minimizes $\|\mathtt{A}\mathbf{f}'\|$ subject to $\|\mathbf{f}'\| = 1$. It turns out not to be possible to do this by linear non-iterative means, chiefly because $\det \mathtt{F}' = 0$ is a cubic, rather than a linear constraint. Nevertheless, it will be seen that a simple iterative method is effective.

An arbitrary singular 3×3 matrix, such as the fundamental matrix $\mathtt{F}$, may be written as a product $\mathtt{F} = \mathtt{M}[\mathbf{e}]_\times$ where $\mathtt{M}$ is a non-singular matrix and $[\mathbf{e}]_\times$ is any skew-symmetric matrix, with $\mathbf{e}$ corresponding to the epipole in the first image.

Suppose we wish to compute the fundamental matrix $\mathtt{F}$ of the form $\mathtt{F} = \mathtt{M}[\mathbf{e}]_\times$ that minimizes the algebraic error $\|\mathtt{A}\mathbf{f}\|$ subject to the condition $\|\mathbf{f}\| = 1$. Let us assume for now that the epipole $\mathbf{e}$ is known. Later we will let $\mathbf{e}$ vary, but for now it is fixed. The equation $\mathtt{F} = \mathtt{M}[\mathbf{e}]_\times$ can be written in terms of the vectors $\mathbf{f}$ and $\mathbf{m}$ comprising the entries of $\mathtt{F}$ and $\mathtt{M}$ as an equation $\mathbf{f} = \mathtt{E}\mathbf{m}$ where $\mathtt{E}$ is a 9×9 matrix. Supposing that $\mathbf{f}$ and $\mathbf{m}$ contain the entries of the corresponding matrices in row-major order, then it can be verified that $\mathtt{E}$ has the form

$$\mathtt{E} = \begin{bmatrix} [\mathbf{e}]_\times & & \\ & [\mathbf{e}]_\times & \\ & & [\mathbf{e}]_\times \end{bmatrix}. \tag{11.4}$$

Since $\mathbf{f} = \mathtt{E}\mathbf{m}$, the minimization problem becomes: [1]

$$\text{Minimize } \|\mathtt{A}\mathtt{E}\mathbf{m}\| \text{ subject to the condition } \|\mathtt{E}\mathbf{m}\| = 1. \tag{11.5}$$

This minimization problem is solved using algorithm A5.6(*p*595). For the purposes of this algorithm one observes that $\text{rank}(\mathtt{E}) = 6$, since each of its diagonal blocks has rank 2.

11.3.1 Iterative estimation

The minimization (11.5) gives a way of computing an algebraic error vector $\mathtt{A}\mathbf{f}$ given a value for the epipole $\mathbf{e}$. This mapping $\mathbf{e} \mapsto \mathtt{A}\mathbf{f}$ is a map from $\mathbb{R}^3$ to $\mathbb{R}^9$. Note that the value of $\mathtt{A}\mathbf{f}$ is unaffected by scaling $\mathbf{e}$. Starting from an estimated value of $\mathbf{e}$ derived as the generator of the right null-space of an initial estimate of $\mathtt{F}$, one may iterate to find the final $\mathtt{F}$ that minimizes algebraic error. The initial estimate of $\mathtt{F}$ may be obtained from the 8-point algorithm, or any other simple algorithm. The complete algorithm for computation of $\mathtt{F}$ is given in algorithm 11.2.

Note the advantage of this method of computing $\mathtt{F}$ is that the iterative part of the algorithm consists of a very small parameter minimization problem, involving the estimation of only three parameters (the homogeneous coordinates of $\mathbf{e}$). Despite this, the algorithm finds the fundamental matrix that minimizes the algebraic error for all matched points. The matched points themselves do not come into the final iterative estimation.

[1] It does not do to minimize $\|\mathtt{A}\mathtt{E}\mathbf{m}\|$ subject to the condition $\|\mathbf{m}\| = 1$, since a solution to this occurs when $\mathbf{m}$ is a unit vector in the right null-space of $\mathtt{E}$. In this case, $\mathtt{E}\mathbf{m} = \mathbf{0}$, and hence $\|\mathtt{A}\mathtt{E}\mathbf{m}\| = 0$.

Objective

Find the fundamental matrix $\mathtt{F}$ that minimizes the algebraic error $\|\mathtt{A}\mathbf{f}\|$ subject to $\|\mathbf{f}\| = 1$ and $\det \mathtt{F} = 0$.

Algorithm

(i) Find a first approximation $\mathtt{F}_0$ for the fundamental matrix using the normalized 8-point algorithm 11.1. Then find the right null-vector $\mathbf{e}_0$ of $\mathtt{F}_0$.

(ii) Starting with the estimate $\mathbf{e}_i = \mathbf{e}_0$ for the epipole, compute the matrix $\mathtt{E}_i$ according to (11.4), then find the vector $\mathbf{f}_i = \mathtt{E}_i\mathbf{m}_i$ that minimizes $\|\mathtt{A}\mathbf{f}_i\|$ subject to $\|\mathbf{f}_i\| = 1$. This is done using algorithm A5.6(*p*595).

(iii) Compute the algebraic error $\boldsymbol{\epsilon}_i = \mathtt{A}\mathbf{f}_i$. Since $\mathbf{f}_i$ and hence $\boldsymbol{\epsilon}_i$ is defined only up to sign, correct the sign of $\boldsymbol{\epsilon}_i$ (multiplying by minus 1 if necessary) so that $\mathbf{e}_i^{\mathsf{T}}\mathbf{e}_{i-1} > 0$ for $i > 0$. This is done to ensure that $\boldsymbol{\epsilon}_i$ varies smoothly as a function of $\mathbf{e}_i$.

(iv) The previous two steps define a mapping $\mathbb{R}^3 \rightarrow \mathbb{R}^9$ mapping $\mathbf{e}_i \mapsto \boldsymbol{\epsilon}_i$. Now use the Levenberg–Marquardt algorithm (section A6.2(*p*600)) to vary $\mathbf{e}_i$ iteratively so as to minimize $\|\boldsymbol{\epsilon}_i\|$.

(v) Upon convergence, $\mathbf{f}_i$ represents the desired fundamental matrix.

Algorithm 11.2. *Computation of* $\mathtt{F}$ *with* $\det \mathtt{F} = 0$ *by iteratively minimizing algebraic error.*

11.4 Geometric distance

This section describes three algorithms which minimize a geometric image distance. The one we recommend, which is the Gold Standard method, unfortunately requires the most effort in implementation. The other algorithms produce extremely good results and are easier to implement, but are not optimal under the assumption that the image errors are Gaussian. Two important issues for each of the algorithms are the intitialization for the non-linear minimization, and the parametrization of the cost function. The algorithms are generally initialized by one of the linear algorithms of the previous section. An alternative, which is used in the automatic algorithm, is to select 7 correspondences and thus generate one or three solutions for $\mathtt{F}$. Various parametrizations are discussed in section 11.4.2. In all cases we recommend that the image points be normalized by a translation and scaling. This normalization does not skew the noise characteristics, so does not interfere with the optimality of the Gold Standard algorithm, which is described next.

11.4.1 The Gold Standard method

The Maximum Likelihood estimate of the fundamental matrix depends on the assumption of an error model. We make the assumption that noise in image point measurements obeys a Gaussian distribution. In that case the ML estimate is the one that minimizes the geometric distance (which is reprojection error)

$$\sum_i d(\mathbf{x}_i, \hat{\mathbf{x}}_i)^2 + d(\mathbf{x}'_i, \hat{\mathbf{x}}'_i)^2 \tag{11.6}$$

where $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ are the measured correspondences, and $\hat{\mathbf{x}}_i$ and $\hat{\mathbf{x}}'_i$ are estimated "true" correspondences that satisfy $\hat{\mathbf{x}}_i'^{\mathsf{T}}\mathtt{F}\hat{\mathbf{x}}_i = 0$ exactly for some rank-2 matrix $\mathtt{F}$, the estimated fundamental matrix.

Objective

Given $n \geq 8$ image point correspondences $\{\mathbf{x}_i \leftrightarrow \mathbf{x}'_i\}$, determine the Maximum Likelihood estimate $\hat{\mathtt{F}}$ of the fundamental matrix.

The MLE involves also solving for a set of subsidiary point correspondences $\{\hat{\mathbf{x}}_i \leftrightarrow \hat{\mathbf{x}}'_i\}$, such that $\hat{\mathbf{x}}_i'^{\mathsf{T}}\hat{\mathtt{F}}\hat{\mathbf{x}}_i = 0$, and which minimizes

$$\sum_i d(\mathbf{x}_i, \hat{\mathbf{x}}_i)^2 + d(\mathbf{x}'_i, \hat{\mathbf{x}}'_i)^2.$$

Algorithm

(i) Compute an initial rank 2 estimate of $\hat{\mathtt{F}}$ using a linear algorithm such as algorithm 11.1.

(ii) Compute an initial estimate of the subsidiary variables $\{\hat{\mathbf{x}}_i, \hat{\mathbf{x}}'_i\}$ as follows:

  (a) Choose camera matrices $\mathtt{P} = [\mathtt{I} \mid \mathbf{0}]$ and $\mathtt{P}' = [[\mathbf{e}']_\times \hat{\mathtt{F}} \mid \mathbf{e}']$, where $\mathbf{e}'$ is obtained from $\hat{\mathtt{F}}$.

  (b) From the correspondence $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ and $\hat{\mathtt{F}}$ determine an estimate of $\widehat{\mathbf{X}}_i$ using the triangulation method of chapter 12.

  (c) The correspondence consistent with $\hat{\mathtt{F}}$ is obtained as $\hat{\mathbf{x}}_i = \mathtt{P}\widehat{\mathbf{X}}_i,\ \hat{\mathbf{x}}'_i = \mathtt{P}'\widehat{\mathbf{X}}_i$.

(iii) Minimize the cost

$$\sum_i d(\mathbf{x}_i, \hat{\mathbf{x}}_i)^2 + d(\mathbf{x}'_i, \hat{\mathbf{x}}'_i)^2$$

over $\hat{\mathtt{F}}$ and $\widehat{\mathbf{X}}_i,\ i = 1, \ldots, n$. The cost is minimized using the Levenberg–Marquardt algorithm over $3n + 12$ variables: $3n$ for the n 3D points $\widehat{\mathbf{X}}_i$, and 12 for the camera matrix $\mathtt{P}' = [\mathtt{M} \mid \mathbf{t}]$, with $\hat{\mathtt{F}} = [\mathbf{t}]_\times \mathtt{M}$, and $\hat{\mathbf{x}}_i = \mathtt{P}\widehat{\mathbf{X}}_i,\ \hat{\mathbf{x}}'_i = \mathtt{P}'\widehat{\mathbf{X}}_i$.

Algorithm 11.3. *The Gold Standard algorithm for estimating* F *from image correspondences*.

This error function may be minimized in the following manner. A pair of camera matrices $\mathtt{P} = [\mathtt{I} \mid \mathbf{0}]$ and $\mathtt{P}' = [\mathtt{M} \mid \mathbf{t}]$ are defined. In addition one defines 3D points $\mathbf{X}_i$. Now letting $\hat{\mathbf{x}}_i = \mathtt{P}\mathbf{X}_i$ and $\hat{\mathbf{x}}'_i = \mathtt{P}'\mathbf{X}_i$, one varies $\mathtt{P}'$ and the points $\mathbf{X}_i$ so as to minimize the error expression. Subsequently $\mathtt{F}$ is computed as $\mathtt{F} = [\mathbf{t}]_\times \mathtt{M}$. The vectors $\hat{\mathbf{x}}_i$ and $\hat{\mathbf{x}}'_i$ will satisfy $\hat{\mathbf{x}}_i'^{\mathsf{T}}\mathtt{F}\hat{\mathbf{x}}_i = 0$. Minimization of the error is carried out using the Levenberg–Marquardt algorithm described in section A6.2(*p*600). An initial estimate of the parameters is computed using the normalized 8-point algorithm, followed by projective reconstruction, as described in chapter 12. Thus, estimation of the fundamental matrix using this method is effectively equivalent to projective reconstruction. The steps of the algorithm are summarized in algorithm 11.3.

It may seem that this method for computing $\mathtt{F}$ will be expensive in computing cost. However, the use of the sparse LM techniques means that it is not much more expensive than other iterative techniques, and details of this are given in section A6.5(*p*609).

11.4.2 Parametrization of rank-2 matrices

The non-linear minimization of the geometric distance cost functions requires a parametrization of the fundamental matrix which enforces the rank 2 property of the matrix. We describe three such parametrizations.

Over-parametrization. One way that we have already seen for parametrizing $\mathtt{F}$ is to write $\mathtt{F} = [\mathbf{t}]_\times \mathtt{M}$, where $\mathtt{M}$ is an arbitrary 3×3 matrix. This ensures that $\mathtt{F}$ is singular, since $[\mathbf{t}]_\times$ is. This way, $\mathtt{F}$ is parametrized by the nine entries of $\mathtt{M}$ and the three entries of $\mathbf{t}$ – a total of 12 parameters, more than the minimum number of parameters, which is 7. In general this should not cause a significant problem.

Epipolar parametrization. An alternative way of parametrizing $\mathtt{F}$ is by specifying the first two columns of $\mathtt{F}$, along with two multipliers α and β such that the third column may be written as a linear combination $\mathbf{f}_3 = \alpha\mathbf{f}_1 + \beta\mathbf{f}_2$. Thus, the fundamental matrix is parametrized as

$$\mathtt{F} = \begin{bmatrix} a & b & \alpha a + \beta b \\ c & d & \alpha c + \beta d \\ e & f & \alpha e + \beta f \end{bmatrix}. \tag{11.7}$$

This has a total of 8 parameters. To achieve a minimum set of parameters, one of the elements, for instance f, may be set to 1. In practice whichever of $a, \ldots, f$ has greatest absolute value is set to 1. This method ensures a singular matrix $\mathtt{F}$, while using the minimum number of parameters. The main disadvantage is that it has a singularity – it does not work when the first two columns of $\mathtt{F}$ are linearly dependent, for then it is not possible to write column 3 in terms of the first two columns. This problem can be significant, since it will occur in the case where the right epipole lies at infinity. For then $\mathtt{F}\mathbf{e} = \mathtt{F}(e_1, e_2, 0)^\mathsf{T} = \mathbf{0}$ and the first two columns are linearly dependent. Nevertheless, this parametrization is widely used and works well if steps are taken to avoid this singularity. Instead of using the first two columns as a basis, another pair of columns can be used, in which case the singularity occurs when the epipole is on one of the coordinate axes. In practice such singularities can be detected during the minimization and the parametrization switched to one of the alternative parametrizations.

Note that $(\alpha, \beta, -1)^\mathsf{T}$ is the right epipole for this fundamental matrix – the coordinates of the epipole occur explicitly in the parametrization. For best results, the parametrization should be chosen so that the largest entry (in absolute value) of the epipole is the one set to 1.

Note how the complete manifold of possible fundamental matrices is not covered by a single parametrization, but rather by a set of minimally parametrized patches. As a path is traced out through the manifold during a parameter minimization procedure, it is necessary to switch from one patch to another as the boundary between patches is crossed. In this case there are actually 18 different parameter patches, depending on which of $a, \ldots, f$ is greatest, and which pair of columns are taken as the basis.

Both epipoles as parameters. The previous parametrization uses one of the epipoles as part of the parametrization. For symmetry one may use both the epipoles as parameters. The resulting form of $\mathtt{F}$ is

$$\mathtt{F} = \begin{bmatrix} a & b & \alpha a + \beta b \\ c & d & \alpha c + \beta d \\ \alpha' a + \beta' c & \alpha' b + \beta' d & \alpha'\alpha a + \alpha'\beta b + \beta'\alpha c + \beta'\beta d \end{bmatrix}. \tag{11.8}$$

The two epipoles are $(\alpha, \beta, -1)^\mathsf{T}$ and $(\alpha', \beta', -1)^\mathsf{T}$. As above, one can set one of a, b, c, d to 1. To avoid singularities, one must switch between different choices of the two rows and two columns to use as the basis. Along with four choices of which of a, b, c, d to set to 1, there are a total of 36 parameter patches used to cover the complete manifold of fundamental matrices.

11.4.3 First-order geometric error (Sampson distance)

The concept of Sampson distance was discussed at length in section 4.2.6(*p*98). Here the Sampson approximation is used in the case of the variety defined by $\mathbf{x}'^\mathsf{T}\mathtt{F}\mathbf{x} = 0$ to provide a first-order approximation to the geometric error.

The general formula for the Sampson cost function is given in (4.13–*p*100). In the case of fundamental matrix estimation, the formula is even simpler, since there is only one equation per point correspondence (see also example 4.2(*p*100)). The partial-derivative matrix $\mathtt{J}$ has only one row, and hence $\mathtt{J}\mathtt{J}^\mathsf{T}$ is a scalar and (4.12–*p*99) becomes

$$\frac{\boldsymbol{\epsilon}^\mathsf{T}\boldsymbol{\epsilon}}{\mathtt{J}\mathtt{J}^\mathsf{T}} = \frac{(\mathbf{x}_i'^\mathsf{T}\mathtt{F}\mathbf{x}_i)^2}{\mathtt{J}\mathtt{J}^\mathsf{T}}.$$

From the definition of $\mathtt{J}$ and the explicit form of $\mathtt{A}_i = \mathbf{x}_i'^\mathsf{T}\mathtt{F}\mathbf{x}_i$ given in the left hand side of (11.2), we obtain

$$\mathtt{J}\mathtt{J}^\mathsf{T} = (\mathtt{F}\mathbf{x}_i)_1^2 + (\mathtt{F}\mathbf{x}_i)_2^2 + (\mathtt{F}^\mathsf{T}\mathbf{x}_i')_1^2 + (\mathtt{F}^\mathsf{T}\mathbf{x}_i')_2^2$$

where for instance $(\mathtt{F}\mathbf{x}_i)_j^2$ represents the square of the j-th entry of the vector $\mathtt{F}\mathbf{x}_i$. Thus, the cost function is

$$\sum_i \frac{(\mathbf{x}_i'^\mathsf{T}\mathtt{F}\mathbf{x}_i)^2}{(\mathtt{F}\mathbf{x}_i)_1^2 + (\mathtt{F}\mathbf{x}_i)_2^2 + (\mathtt{F}^\mathsf{T}\mathbf{x}_i')_1^2 + (\mathtt{F}^\mathsf{T}\mathbf{x}_i')_2^2}. \tag{11.9}$$

This gives a first-order approximation to geometric error, which may be expected to give good results if higher order terms are small in comparison to the first. The approximation has been used successfully in estimation algorithms by [Torr-97, Torr-98, Zhang-98]. Note that this approximation is undefined at the point in $\mathbb{R}^4$ determined by the two epipoles, as here $\mathtt{J}\mathtt{J}^\mathsf{T}$ is zero. This point should be avoided in any numerical implementation.

The key advantage of approximating the geometric error in this way is that the resulting cost function only involves the parameters of $\mathtt{F}$. This means that to first-order the Gold Standard cost function (11.6) is minimized *without* introducing a set of subsidiary variables, namely the coordinates of the n space points $\mathbf{X}_i$. Consequently a minimization problem with $7 + 3n$ degrees of freedom is reduced to one with only 7 degrees of freedom.

Symmetric epipolar distance. Equation (11.9) is similar in form to another cost function

$$\sum_i \quad d(\mathbf{x}_i', \mathtt{F}\mathbf{x}_i)^2 + d(\mathbf{x}_i, \mathtt{F}^\mathsf{T}\mathbf{x}_i')^2$$

$$= \sum_i (\mathbf{x}_i'^{\mathsf{T}} \mathtt{F} \mathbf{x}_i)^2 \left(\frac{1}{(\mathtt{F}\mathbf{x}_i)_1^2 + (\mathtt{F}\mathbf{x}_i)_2^2} + \frac{1}{(\mathtt{F}^{\mathsf{T}}\mathbf{x}_i')_1^2 + (\mathtt{F}^{\mathsf{T}}\mathbf{x}_i')_2^2} \right) \quad (11.10)$$

which minimizes the distance of a point from its projected epipolar line, computed in each of the images. However, this cost function seems to give slightly inferior results to (11.9) (see [Zhang-98]), and hence is not discussed further.

11.5 Experimental evaluation of the algorithms

Three of the algorithms of the previous sections are now compared by estimating F from point correspondences for a number of image pairs. The algorithms are:

(i) The normalized 8-point algorithm (algorithm 11.1).
(ii) Minimization of algebraic error whilst imposing the singularity constraint (algorithm 11.2).
(iii) The Gold Standard geometric algorithm (algorithm 11.3).

The experimental procedure was as follows. For each pair of images, a number n of matched points were chosen randomly from the matches and the fundamental matrix estimated and residual error (see below) computed. This experiment was repeated 100 times for each value of n and each pair of images, and the average residual error plotted against n. This gives an idea of how the different algorithms behave as the number of points is increased. The number of points used, n, ranged from 8 up to three-quarters of the total number of matched points.

Residual error

The error is defined as

$$\frac{1}{N} \sum_i^N d(\mathbf{x}_i', \mathtt{F}\mathbf{x}_i)^2 + d(\mathbf{x}_i, \mathtt{F}^{\mathsf{T}}\mathbf{x}_i')^2$$

where $d(\mathbf{x}, \mathbf{l})$ here is the distance (in pixels) between a point $\mathbf{x}$ and a line $\mathbf{l}$. The error is the squared distance between a point's epipolar line and the matching point in the other image (computed for both points of the match), averaged over all N matches. Note the error is evaluated over *all* N matched points, and not just the n matches used to compute F. The residual error corresponds to the epipolar distance defined in (11.10). Note that this particular error is *not* minimized directly by any of the algorithms evaluated here.

The various algorithms were tried with 5 different pairs of images. The images are presented in figure 11.2 and show the diversity of image types, and placement of the epipoles. A few of the epipolar lines are shown in the images. The intersection of the pencil of lines is the epipole. There was a wide variation in the accuracy of the matched points for the different images, though mismatches were removed in a pre-processing step.

Results. The results of these experiments are shown and explained in figure 11.3. They show that minimizing algebraic error gives essentially indistinguishable results from minimizing geometric error.

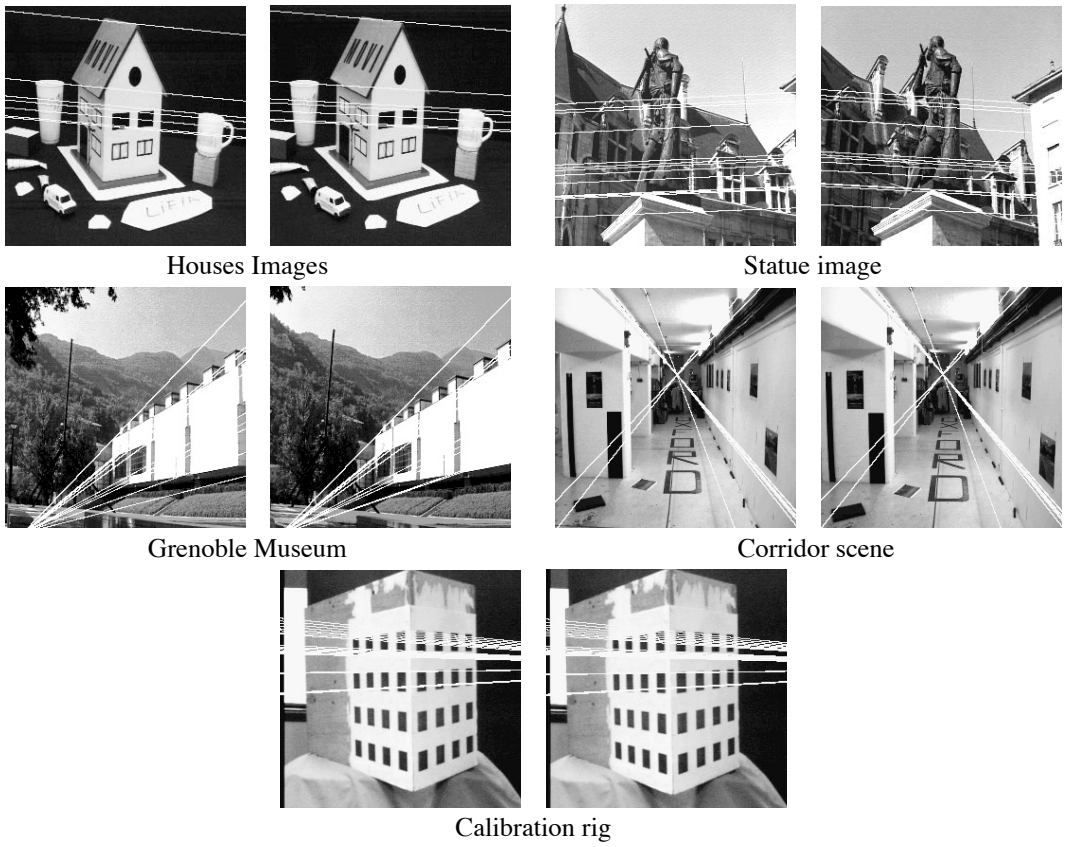


Fig. 11.2. *Image pairs used for the algorithm comparison. In the top two the epipoles are far from the image centres. In the middle two the epipoles are close (Grenoble) and in the image (Corridor). For the calibration images the matched points are known extremely accurately.*

11.5.1 Recommendations

Several methods of computing the fundamental matrix have been discussed in this chapter, and some pointers on which method to use are perhaps desirable. Briefly, these are our recommendations:

- Do not use the unnormalized 8-point algorithm.
- For a quick method, easy to implement, use the normalized 8-point algorithm 11.1. This often gives adequate results, and is ideal as a first step in other algorithms.
- If more accuracy is desired, use the algebraic minimization method, either with or without iteration on the position of the epipole.
- As an alternative that gives excellent results, use an iterative-minimization method that minimizes the Sampson cost function (11.9). This and the iterative algebraic method give similar results.
- To be certain of getting the best results, if Gaussian noise is a viable assumption, implement the Gold Standard algorithm.

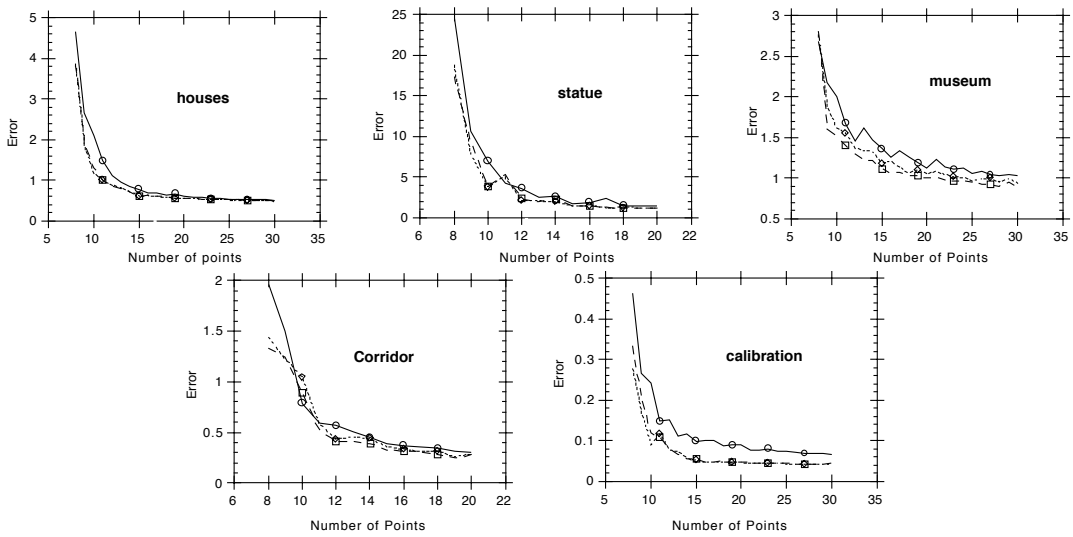


Fig. 11.3. **Results of the experimental evaluation of the algorithms.** *In each case, three methods of computing* F *are compared. Residual error is plotted against the number of points used to compute* F*. In each graph, the top (solid line) shows the results of the normalized 8-point algorithm. Also shown are the results of minimizing geometric error (long dashed line) and iteratively minimizing algebraic error subject to the determinant constraint (short dashed line). In most cases, the result of iteratively minimizing algebraic error is almost indistinguishable from minimizing geometric error. Both are noticeably better than the non-iterative normalized 8-point algorithm, though that algorithm also gives good results.*

11.6 Automatic computation of F

This section describes an algorithm to compute the epipolar geometry between two images automatically. The input to the algorithm is simply the pair of images, with no other *a priori* information required; and the output is the estimated fundamental matrix together with a set of interest points in correspondence.

The algorithm uses RANSAC as a search engine in a similar manner to its use in the automatic computation of a homography described in section 4.8(*p*123). The ideas and details of the algorithm are given there, and are not repeated here. The method is summarized in algorithm 11.4, with an example of its use shown in figure 11.4.

A few remarks on the method:

(i) **The RANSAC sample.** Only 7 point correspondences are used to estimate F. This has the advantage that a rank 2 matrix is produced, and it is not necessary to coerce the matrix to rank 2 as in the linear algorithms. A second reason for using 7 correspondences, rather than 8 say with a linear algorithm, is that the number of samples that must be tried in order to ensure a high probability of no outliers is exponential in the size of the sample set. For example, from table 4.3-(*p*119) for a 99% confidence of no outliers (when drawing from a set containing 50% outliers) twice as many samples are required for 8 correspondences as for 7. The slight disadvantage in using 7 correspondences is that it may result in 3 real solutions for F, and all 3 must be tested for support.

Objective Compute the fundamental matrix between two images.

Algorithm

(i) **Interest points:** Compute interest points in each image.
(ii) **Putative correspondences:** Compute a set of interest point matches based on proximity and similarity of their intensity neighbourhood.
(iii) **RANSAC robust estimation:** Repeat for N samples, where N is determined adaptively as in algorithm 4.5(*p*121):
  (a) Select a random sample of 7 correspondences and compute the fundamental matrix F as described in section 11.1.2. There will be one or three real solutions.
  (b) Calculate the distance $d_\perp$ for each putative correspondence.
  (c) Compute the number of inliers consistent with F by the number of correspondences for which $d_\perp < t$ pixels.
  (d) If there are three real solutions for F the number of inliers is computed for each solution, and the solution with most inliers retained.

  Choose the F with the largest number of inliers. In the case of ties choose the solution that has the lowest standard deviation of inliers.
(iv) **Non-linear estimation:** re-estimate F from all correspondences classified as inliers by minimizing a cost function, e.g. (11.6), using the Levenberg–Marquardt algorithm of section A6.2(*p*600).
(v) **Guided matching:** Further interest point correspondences are now determined using the estimated F to define a search strip about the epipolar line.

The last two steps can be iterated until the number of correspondences is stable.

Algorithm 11.4. *Algorithm to automatically estimate the fundamental matrix between two images using RANSAC.*

(ii) **The distance measure.** Given a current estimate of F (from the RANSAC sample) the distance $d_\perp$ measures how closely a matched pair of points satisfies the epipolar geometry. There are two clear choices for $d_\perp$: reprojection error, i.e. the distance minimized in the cost function (11.6) (the value may be obtained using the triangulation algorithm of section 12.5); or the Sampson approximation to reprojection error ($d_\perp^2$ is given by (11.9)). If the Sampson approximation is used, then the Sampson cost function should be used to iteratively estimate F. Otherwise distances used in RANSAC and elsewhere in the algorithm will be inconsistent.

(iii) **Guided matching.** The current estimate of F defines a search band in the second image around the epipolar line $\mathtt{F}\mathbf{x}$ of $\mathbf{x}$. For each corner $\mathbf{x}$ a match is sought within this band. Since the search area is restricted a weaker similarity threshold can be employed, and it is not necessary to enforce a "winner takes all" scheme.

(iv) **Implementation and run details.** For the example of figure 11.4, the search window was ± 300 pixels. The inlier threshold was $t = 1.25$ pixels. A total of 407 samples were required. The RMS pixel error after RANSAC was 0.34 (for 99 correspondences), and after MLE and guided matching it was 0.33 (for 157 correspondences). The guided matching MLE required 10 iterations of the Levenberg–Marquardt algorithm.

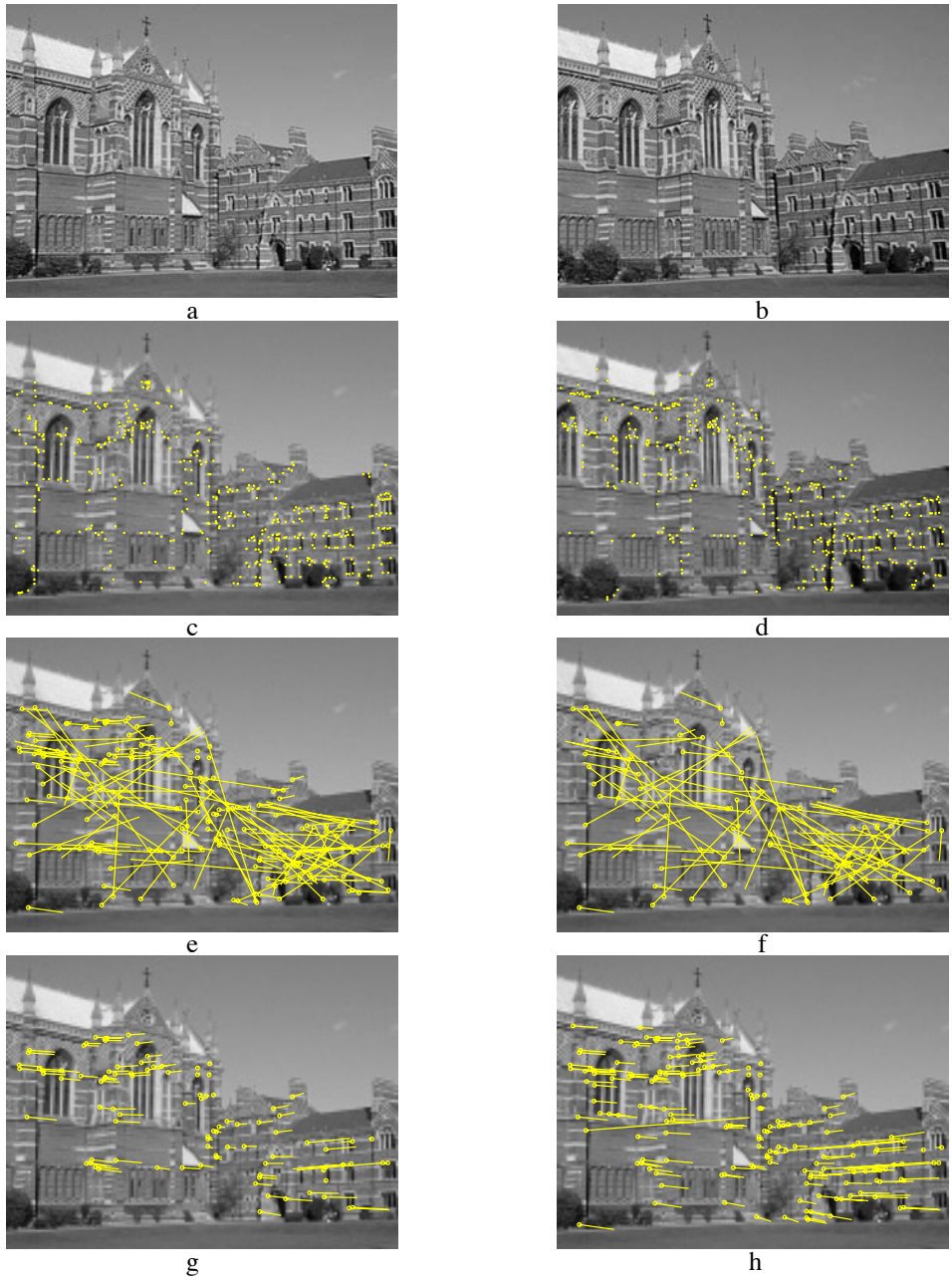


Fig. 11.4. **Automatic computation of the fundamental matrix between two images using RANSAC.** *(a) (b) left and right images of Keble College, Oxford. The motion between views is a translation and rotation. The images are* 640×480 *pixels. (c) (d) detected corners superimposed on the images. There are approximately 500 corners on each image. The following results are superimposed on the left image: (e) 188 putative matches shown by the line linking corners, note the clear mismatches; (f) outliers – 89 of the putative matches. (g) inliers – 99 correspondences consistent with the estimated* F*; (h) final set of 157 correspondences after guided matching and MLE. There are still a few mismatches evident, e.g. the long line on the left.*

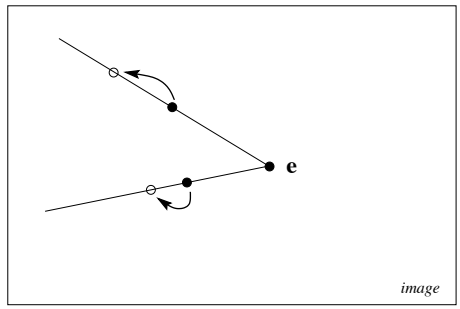


Fig. 11.5. *For a pure translation the epipole can be estimated from the image motion of two points.*

11.7 Special cases of F-computation

Certain special cases of motion, or partially known camera calibration, allow computation of the fundamental matrix to be simplified. In each case the number of degrees of freedom of the fundamental matrix is less than the 7 of general motion. We give three examples.

11.7.1 Pure translational motion

This is the simplest possible case. The matrix can be estimated linearly whilst simultaneously imposing the constraints that the matrix must satisfy, namely that it is skew-symmetric (see section 9.3.1(*p*247)), and thus has the required rank of 2. In this case $\mathtt{F} = [\mathbf{e}']_\times$, and has two degrees of freedom. It may be parametrized by the three entries of $\mathbf{e}'$.

Each point correspondence provides one linear constraint on the homogeneous parameters, as is clear from figure 11.5. The matrix can be computed uniquely from two point correspondences.

Note, in the general motion case if all 3D points are coplanar, which is a structure degeneracy (see section 11.9), the fundamental matrix cannot be determined uniquely from image correspondences. However, for pure translational motion this is not a problem (two 3D points are always coplanar). The only degeneracy is if the two 3D points are coplanar with both camera centres.

This special form also simplifies the Gold Standard estimation, and correspondingly triangulation for structure recovery. The Gold Standard estimation of the epipole from point correspondences under pure translation is identical to the estimation of a vanishing point given the end points of a set of imaged parallel lines, see section 8.6.1(*p*213).

11.7.2 Planar motion

In the case of planar motion, described in section 9.3.2(*p*250), we require that the symmetric part of F has rank 2, in addition to the standard rank 2 condition for the full matrix. It can be verified that the parametrization of (9.8–*p*252), namely $\mathtt{F} = [\mathbf{e}']_\times [\mathbf{l}_s]_\times [\mathbf{e}]_\times$, satisfies both these conditions. If unconstrained 3-vectors are used to represent $\mathbf{e}', \mathbf{l}_s$ and $\mathbf{e}$ then 9 parameters are used, whereas the fundamental matrix for planar motion has only 6 degrees of freedom. As usual this over-parametrization is not a problem.

An alternative parametrization with similar properties is

$$\mathtt{F} = \alpha[\mathbf{x_a}]_\times + \beta\left(\mathbf{l}_s\mathbf{l}_h^\mathsf{T} + \mathbf{l}_h\mathbf{l}_s^\mathsf{T}\right) \quad \text{with} \quad \mathbf{x}_\mathtt{a}^\mathsf{T}\mathbf{l}_h = 0$$

where α and β are scalars, and the meaning of the 3-vectors $\mathbf{x_a}, \mathbf{l}_s$ and $\mathbf{l}_h$ is evident from figure 9.11(*p*253)(a).

11.7.3 The calibrated case

In the case of calibrated cameras normalized image coordinates may be used, and the essential matrix E computed instead of the fundamental matrix. As with the fundamental matrix, the essential matrix may be computed using linear techniques from 8 points or more, since corresponding points satisfy the defining equation $\mathbf{x}_i'^\mathsf{T}\mathtt{E}\mathbf{x}_i = 0$.

Where the method differs from the computation of the fundamental matrix is in the enforcement of the constraints. For, whereas the fundamental matrix satisfies $\det \mathtt{F} = 0$, the essential matrix satisfies the additional condition that its two singular values are equal. This constraint may be handled by the following result, which is offered here without proof.

Result 11.1. *Let* E *be a* 3×3 *matrix with SVD given by* $\mathtt{E} = \mathtt{U}\mathtt{D}\mathtt{V}^\mathsf{T}$*, where* $\mathtt{D} = \mathrm{diag}(a, b, c)$ *with* $a \geq b \geq c$*. Then the closest essential matrix to* E *in Frobenius norm is given by* $\hat{\mathtt{E}} = \mathtt{U}\hat{\mathtt{D}}\mathtt{V}^\mathsf{T}$*, where* $\hat{\mathtt{D}} = \mathrm{diag}((a+b)/2, (a+b)/2, 0)$.

If the goal is to compute the two normalized camera matrices P and P′ as part of a reconstruction process, then it is not actually necessary to compute $\hat{\mathtt{E}}$ by multiplying out $\hat{\mathtt{E}} = \mathtt{U}\hat{\mathtt{D}}\mathtt{V}^\mathsf{T}$. Matrix P′ can be computed directly from the SVD according to result 9.19-(*p*259). The choice between the four solutions for P′ is determined by the consideration that the visible points must lie in front of the two cameras, as explained in section 9.6.3-(*p*259).

11.8 Correspondence of other entities

So far in this chapter only point correspondences have been employed, and the question naturally arises: can F be computed from the correspondence of image entities other than points? The answer is yes, but not from all types of entities. We will now discuss some common examples.

Lines. The correspondence of image lines between views places *no* constraint at all on F. Here a line is an infinite line, not a line segment. Consider the case of corresponding image points: the points in each image back-project to rays, one through each camera centre, and these rays intersect at the 3-space point. Now in general two lines in 3-space are skew (i.e. they do not intersect); so the condition that the rays intersect places a constraint on the epipolar geometry. In contrast in the case of corresponding image lines, the back-projection is a plane from each view. However, two planes in 3-space always intersect so there is no constraint on the epipolar geometry (there is a constraint in the case of 3-views).

In the case of parallel lines, the correspondence of vanishing points does provide a

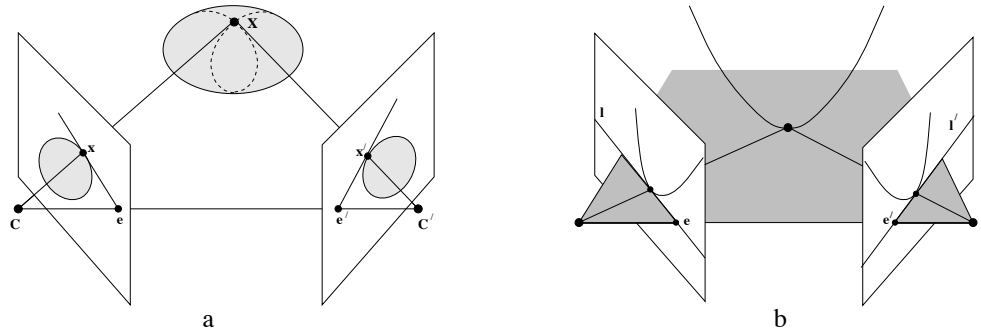


Fig. 11.6. **Epipolar tangency.** *(a) for a surface; (b) for a space curve – figure after Porrill and Pollard [Porrill-91]. In (a) the epipolar plane* $\mathbf{CC'X}$ *is tangent to the surface at* $\mathbf{X}$*. The imaged outline is tangent to the epipolar lines at* $\mathbf{x}$ *and* $\mathbf{x}'$ *in the two views. The dashed curves on the surface are the contour generators. In (b) the epipolar plane is tangent to the space curve. The corresponding epipolar lines* $\mathbf{l} \leftrightarrow \mathbf{l}'$ *are tangent to the imaged curve.*

constraint on $\mathtt{F}$. However, a vanishing point has the same status as any finite point, i.e. it provides one constraint.

Space curves and surfaces. As illustrated in figure 11.6, at points at which the epipolar plane is tangent to a space curve the imaged curve is tangent to the corresponding epipolar lines. This provides a constraint on the 2 view geometry, i.e. if an epipolar line is tangent to an imaged curve in one view, then the corresponding epipolar line must be tangent to the imaged curve in the other view. Similarly, in the case of surfaces, at points at which the epipolar plane is tangent to the surface the imaged outline is tangent to the corresponding epipolar lines. Epipolar tangent points act effectively as point correspondences and may be included in estimation algorithms as described by [Porrill-91].

Particularly important cases are those of conics and quadrics which are algebraic objects and so algebraic solutions can be developed. Examples are given in the notes and exercises at the end of this chapter.

11.9 Degeneracies

A set of correspondences $\{\mathbf{x}_i \leftrightarrow \mathbf{x}'_i, i = 1, \ldots, n\}$ is geometrically degenerate with respect to $\mathtt{F}$ if it fails to uniquely define the epipolar geometry, or equivalently if there exist linearly independent rank-2 matrices, $\mathtt{F}_j, j = 1, 2$, such that

$$\mathbf{x}_i'^{\mathsf{T}} \mathtt{F}_1 \mathbf{x}_i^{\mathsf{T}} = 0 \quad \text{and} \quad \mathbf{x}_i'^{\mathsf{T}} \mathtt{F}_2 \mathbf{x}_i = 0 \qquad (1 \leq i \leq n) \ .$$

The subject of degeneracy is investigated in detail in chapter 22. However, a brief preview is given now for the two important cases of scene points on a ruled quadric, or on a plane.

Provided the two camera centres are not coincident the epipolar geometry *is* uniquely defined. It can always be computed from the camera matrices $\mathtt{P}, \mathtt{P}'$ as in (9.1–*p*244) for example. What is at issue here are configurations where the epipolar geometry cannot be estimated from point correspondences. An awareness of the degeneracies of

$\dim(N) = 1$: Unique solution – no degeneracy.

Arises from $n \geq 8$ point correspondences in general position. If $n > 8$ then the point correspondences must be perfect (i.e. noise-free).

$\dim(N) = 2$: 1 or 3 solutions.

Arises in the case of seven point correspondences, and also in the case of $n > 7$ perfect point correspondences where the 3D points and camera centres lie on a ruled quadric referred to as a critical surface. The quadric may be non-degenerate (a hyperboloid of one sheet) or degenerate.

$\dim(N) = 3$: Two-parameter family of solutions.

Arises if $n \geq 6$ perfect point correspondences are related by a homography, $\mathbf{x}'_i = \mathtt{H}\mathbf{x}_i$.

- Rotation about the camera centre (a degenerate motion).
- All world points on a plane (a degenerate structure).

Table 11.1. *Degeneracies in estimating* F *from point correspondences, classified by the dimension of the null-space* N *of* A *in (11.3–*p*279).*

estimation algorithms is important because configurations "close to" degenerate ones are likely to lead to a numerically ill-conditioned estimation. The degeneracies are summarized in table 11.1.

11.9.1 Points on a ruled quadric

It will be shown in chapter 22 that degeneracy occurs if both camera centres and all the 3D points lie on a (ruled) quadric surface referred to as the *critical surface* [Maybank-93]. A ruled quadric may be non-degenerate (a hyperboloid of one sheet – a cooling tower) or degenerate (for instance two planes, cones, and cylinders) – see section 3.2.4(*p*74); but a critical surface cannot be an ellipsoid or hyperboloid of two sheets. For a critical surface configuration there are three possible fundamental matrices.

Note that in the case of just 7 point correspondences, together with the two camera centres there are 9 points in total. A general quadric has 9 degrees of freedom, and one may always construct a quadric through 9 points. In the case where this quadric is a ruled quadric it will be a critical surface, and there will be three possible solutions for F. The case where the quadric is not ruled corresponds to the case where there is only one real solution for F.

11.9.2 Points on a plane

An important degeneracy is when all the points lie in a plane. In this case, all the points plus the two camera centres lie on a ruled quadric surface, namely the degenerate quadric consisting of two planes – the plane through the points, plus a plane passing through the two camera centres.

Two views of a planar set of points are related via a 2D projective transformation H. Thus, suppose that a set of correspondences $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ is given for which $\mathbf{x}'_i = \mathtt{H}\mathbf{x}_i$. Any number of points $\mathbf{x}_i$ and the corresponding points $\mathbf{x}'_i = \mathtt{H}\mathbf{x}_i$ may be given.

The fundamental matrix corresponding to the pair of cameras satisfies the equation $\mathbf{x}_i'^{\mathsf{T}}\mathtt{F}\mathbf{x}_i = \mathbf{x}_i'^{\mathsf{T}}(\mathtt{F}\mathtt{H}^{-1})\mathbf{x}_i' = 0$. This set of equations is satisfied whenever $\mathtt{F}\mathtt{H}^{-1}$ is skew-symmetric. Thus, the solution for $\mathtt{F}$ is any matrix of the form $\mathtt{F} = \mathtt{S}\mathtt{H}$, where $\mathtt{S}$ is skew-symmetric. Now, a 3×3 skew-symmetric matrix $\mathtt{S}$ may be written in the form $\mathtt{S} = [\mathbf{t}]_\times$, for any 3-vector $\mathbf{t}$. Thus, $\mathtt{S}$ has three degrees of freedom, and consequently so does $\mathtt{F}$. More precisely, the correspondences $\mathbf{x}_i \leftrightarrow \mathbf{x}_i'$ lead to a three-parameter family of possible fundamental matrices $\mathtt{F}$ (note, one of the parameters accounts for scaling the matrix so there is only a *two*-parameter family of homogeneous matrices). The equation matrix $\mathtt{A}$ derived from the set of correspondences must therefore have rank no greater than 6.

From the decomposition of $\mathtt{F} = \mathtt{S}\mathtt{H}$, it follows from result 9.9(*p*254) that the pair of camera matrices $[\mathtt{I} \mid \mathbf{0}]$ and $[\mathtt{H} \mid \mathbf{t}]$ correspond to the fundamental matrix $\mathtt{F}$. Here, the vector $\mathbf{t}$ may take on any value. If point $\mathbf{x}_i = (x_i, y_i, 1)^{\mathsf{T}}$ and $\mathbf{x}_i' = \mathtt{H}\mathbf{x}_i$, then one verifies that the point $\mathbf{X}_i = (x, y, 1, 0)^{\mathsf{T}}$ maps to $\mathbf{x}_i$ and $\mathbf{x}_i'$ through the two cameras. Thus, the points $\mathbf{X}_i$ constitute a reconstruction of the scene.

11.9.3 No translation

If the two camera centres are coincident then the epipolar geometry is not defined. In addition, formulae such as result 9.9(*p*254) give a value of $\mathbf{0}$ for the fundamental matrix. In this case the two images are related by a 2D homography (see section 8.4.2(*p*204)).

If one attempts to find the fundamental matrix then, as shown above, there will be at least a 2-parameter family of solutions for $\mathtt{F}$. Even if the camera motion involves no translation, then a method such as the 8-point algorithm used to compute the fundamental matrix will still produce a matrix $\mathtt{F}$ satisfying $\mathbf{x}_i'^{\mathsf{T}}\mathtt{F}\mathbf{x}_i = 0$, where $\mathtt{F}$ has the form $\mathtt{F} = \mathtt{S}\mathtt{H}$, $\mathtt{H}$ is the homography relating the points, and $\mathtt{S}$ is an essentially arbitrary skew-symmetric matrix. Points $\mathbf{x}_i$ and $\mathbf{x}_i'$ related by $\mathtt{H}$ will satisfy this relationship.

11.10 A geometric interpretation of F-computation

The estimation of $\mathtt{F}$ from a set of image correspondences $\{\mathbf{x}_i \leftrightarrow \mathbf{x}_i'\}$ has many similarities with the problem of estimating a conic from a set of 2D points $\{x_i, y_i\}$ (or a quadric from a set of 3D points).

The equation $\mathbf{x}'^{\mathsf{T}}\mathtt{F}\mathbf{x} = 0$ is a single constraint in x, y, x', y' and so defines a surface (variety) $\mathcal{V}$ of codimension 1 (dimension 3) in $\mathbb{R}^4$. The surface is a quadric because the equation is quadratic in the coordinates x, y, x', y' of $\mathbb{R}^4$. There is a natural mapping from projective 3-space to the variety $\mathcal{V}$ that takes any 3D point to the quadruple $(x, y, x', y')^{\mathsf{T}}$ of the corresponding image points in the two views. The quadric form is evident if $\mathbf{x}'^{\mathsf{T}}\mathtt{F}\mathbf{x} = 0$ is rewritten as

$$\begin{pmatrix} x & y & x' & y' & 1 \end{pmatrix} \begin{bmatrix} 0 & 0 & f_{11} & f_{21} & f_{31} \\ 0 & 0 & f_{12} & f_{22} & f_{32} \\ f_{11} & f_{12} & 0 & 0 & f_{13} \\ f_{21} & f_{22} & 0 & 0 & f_{23} \\ f_{31} & f_{32} & f_{13} & f_{23} & 2f_{33} \end{bmatrix} \begin{pmatrix} x \\ y \\ x' \\ y' \\ 1 \end{pmatrix} = 0 \ .$$

The case of conic fitting is a good (lower-dimensional) model of $\mathtt{F}$ estimation. To

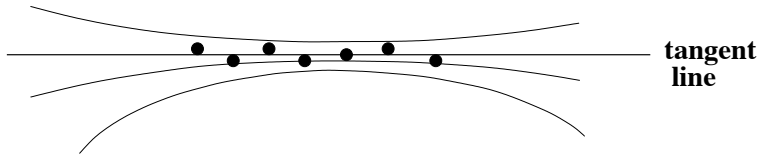


Fig. 11.7. *Estimating a conic from point data (shown as •) may be poorly conditioned. All of the conics shown have residuals within the point error distribution. However, even though there is ambiguity in the estimated conic, the tangent line is well defined, and can be computed from the points.*

bring out the analogy between the two estimation problems: a point (x_i, y_i) places one constraint on the 5 degrees of freedom of a conic as described in section 2.2.3(*p*30):

$$ax_i^2 + bx_iy_i + cy_i^2 + dx_i + ey_i + f = 0.$$

Similarly, a point correspondence (x_i, y_i, x_i', y_i') places one constraint on the (8) degrees of freedom of $\mathtt{F}$ as (11.2–*p*279):

$$x_i'x_if_{11} + x_i'y_if_{12} + x_i'f_{13} + y_i'x_if_{21} + y_i'y_if_{22} + y_i'f_{23} + x_if_{31} + y_if_{32} + f_{33} = 0.$$

It is not quite an exact analogy, since the defining relationship expressed by the fundamental matrix is bilinear in the two sets of indices, as is also evident from the zeros in the quadric matrix above, whereas in the case of a conic section the equation is an arbitrary quadratic. Also the surface defined by $\mathtt{F}$ must satisfy an additional constraint arising from $\det(\mathtt{F}) = 0$, and there is no such constraint in the conic fitting analogue.

The problems of extrapolation when data has only been fitted to a small section of a conic are well known, and similar issues arise in fitting the fundamental matrix to data. Indeed, there are cases where the data is sufficient to determine an accurate tangent line to the conic, but insufficient to determine the conic itself, see figure 11.7. In the case of the fundamental matrix the tangent plane to the quadric in $\mathbb{R}^4$ is the affine fundamental matrix (chapter 14), and this approximation may be fitted when perspective effects are small.

11.11 The envelope of epipolar lines

One of the uses of the fundamental matrix is to determine epipolar lines in a second image corresponding to points in a first image. For instance, if one is seeking matched points between two images, the match of a given point $\mathbf{x}$ in the first image may be found by searching along the epipolar line $\mathtt{F}\mathbf{x}$ in the second image. In the presence of noise, of course, the matching point will not lie precisely on the line $\mathtt{F}\mathbf{x}$ because the fundamental matrix will be known only within certain bounds, expressed by its covariance matrix. In general, instead of searching along the epipolar line only, it will be necessary to search in a region on either side of the line $\mathtt{F}\mathbf{x}$. We will now consider how the covariance matrix of the fundamental matrix may be used to determine the region in which to search.

Let $\mathbf{x}$ be a point and $\mathtt{F}$ be a fundamental matrix for which one has computed a covariance matrix $\Sigma_{\mathtt{F}}$. The point $\mathbf{x}$ corresponds to an epipolar line $\mathbf{l} = \mathtt{F}\mathbf{x}$, and one

may transfer the covariance matrix $\Sigma_{\mathtt{F}}$ to a covariance matrix $\Sigma_{\mathtt{l}}$ according to result 5.6-(p139). Also by result 5.6(p139), the mean value of the epipolar line is given by $\bar{\mathtt{l}} = \bar{\mathtt{F}}\mathbf{x}$. To avoid singular cases, the vector $\mathtt{l}$ representing an epipolar line is normalized so that $\|\mathtt{l}\| = 1$. Then the mapping $\mathbf{x} \mapsto \mathtt{l}$ is given by $\mathtt{l} = (\mathtt{F}\mathbf{x})/\|\mathtt{F}\mathbf{x}\|$. If $\mathtt{J}$ is the Jacobian matrix of this mapping with respect to the entries of $\mathtt{F}$, then $\mathtt{J}$ is a 3×9 matrix, and $\Sigma_{\mathtt{l}} = \mathtt{J}\Sigma_{\mathtt{F}}\mathtt{J}^{\mathsf{T}}$.

Though the constraint $\|\mathtt{l}\| = 1$ is the most convenient constraint, the following analysis applies for any constraint used to confine the vector representing the epipolar line to vary on a 2-dimensional surface in $\mathbb{R}^3$. In this case, the covariance matrix $\Sigma_{\mathtt{l}}$ is singular, having rank 2, since no variation is allowed in the direction normal to the constraint surface. For a particular instance of $\mathtt{l}$, the deviation from the mean, $\bar{\mathtt{l}} - \mathtt{l}$, must be along the constraint surface, and hence (in the linear approximation) perpendicular to the null-space of $\Sigma_{\mathtt{l}}$.

For the remainder of this derivation, $\bar{\mathtt{l}}$, the vector representing the mean epipolar line, will be denoted by $\mathbf{m}$, so as to avoid confusing notation. Now, assuming a Gaussian distribution for the vectors $\mathtt{l}$ representing the epipolar line, the set of all lines having a given likelihood is given by the equation

$$(\mathbf{l} - \mathbf{m})^{\mathsf{T}}\Sigma_{\mathtt{l}}^{+}(\mathbf{l} - \mathbf{m}) = k^2 \tag{11.11}$$

where k is some constant. To analyze this further, we apply an orthogonal change of coordinates such that $\Sigma_{\mathtt{l}}$ becomes diagonal. Thus, one may write

$$\mathtt{U}\Sigma_{\mathtt{l}}\mathtt{U}^{\mathsf{T}} = \Sigma'_{\mathtt{l}} = \begin{bmatrix} \tilde{\Sigma}'_{\mathtt{l}} & \mathbf{0} \\ \mathbf{0}^{\mathsf{T}} & 0 \end{bmatrix}$$

where $\tilde{\Sigma}'_{\mathtt{l}}$ is a 2×2 non-singular diagonal matrix. Applying the same transformation to the lines, one defines 2-vectors $\mathbf{m}' = \mathtt{U}\mathbf{m}$ and $\mathbf{l}' = \mathtt{U}\mathbf{l}$. Since $\mathbf{l}' - \mathbf{m}'$ is orthogonal to the null-space $(0, 0, 1)^{\mathsf{T}}$ of $\Sigma'_{\mathtt{l}}$, both $\mathbf{m}'$ and $\mathbf{l}'$ have the same third coordinate. By multiplying $\mathtt{U}$ by a constant as necessary, one may assume that this coordinate is 1. Thus we may write $\mathbf{l}' = (\tilde{\mathbf{l}}'^{\mathsf{T}}, 1)^{\mathsf{T}}$ and $\mathbf{m}' = (\tilde{\mathbf{m}}'^{\mathsf{T}}, 1)^{\mathsf{T}}$ for certain 2-vectors $\tilde{\mathbf{l}}'$ and $\tilde{\mathbf{m}}'$. Then, one verifies that

$$\begin{aligned} k^2 &= (\mathbf{l} - \mathbf{m})^{\mathsf{T}}\Sigma_{\mathtt{l}}^{+}(\mathbf{l} - \mathbf{m}) \\ &= (\mathbf{l}' - \mathbf{m}')^{\mathsf{T}}\Sigma_{\mathtt{l}}'^{+}(\mathbf{l}' - \mathbf{m}') \\ &= (\tilde{\mathbf{l}}' - \tilde{\mathbf{m}}')^{\mathsf{T}}\tilde{\Sigma}_{\mathtt{l}}'^{-1}(\tilde{\mathbf{l}}' - \tilde{\mathbf{m}}'). \end{aligned}$$

This equation expands out to

$$\tilde{\mathbf{l}}'^{\mathsf{T}}\tilde{\Sigma}_{\mathtt{l}}'^{-1}\tilde{\mathbf{l}}' - \tilde{\mathbf{m}}'^{\mathsf{T}}\tilde{\Sigma}_{\mathtt{l}}'^{-1}\tilde{\mathbf{l}}' - \tilde{\mathbf{l}}'^{\mathsf{T}}\tilde{\Sigma}_{\mathtt{l}}'^{-1}\tilde{\mathbf{m}}' + \tilde{\mathbf{m}}'^{\mathsf{T}}\tilde{\Sigma}_{\mathtt{l}}'^{-1}\tilde{\mathbf{m}}' - k^2 = 0$$

which may be written as

$$(\tilde{\mathbf{l}}'^{\mathsf{T}} \;\; 1) \begin{bmatrix} \tilde{\Sigma}_{\mathtt{l}}'^{-1} & -\tilde{\Sigma}_{\mathtt{l}}'^{-1}\tilde{\mathbf{m}}' \\ -\tilde{\mathbf{m}}'^{\mathsf{T}}\tilde{\Sigma}_{\mathtt{l}}'^{-1} & \tilde{\mathbf{m}}'^{\mathsf{T}}\tilde{\Sigma}_{\mathtt{l}}'^{-1}\tilde{\mathbf{m}}' - k^2 \end{bmatrix} \begin{pmatrix} \tilde{\mathbf{l}}' \\ 1 \end{pmatrix} = 0$$

or equivalently (as one may verify)

$$(\tilde{\mathbf{l}}'^{\mathsf{T}} \;\; 1) \begin{bmatrix} \tilde{\mathbf{m}}'\tilde{\mathbf{m}}'^{\mathsf{T}} - k^2\tilde{\Sigma}'_{\mathbf{l}} & \tilde{\mathbf{m}}' \\ \tilde{\mathbf{m}}'^{\mathsf{T}} & 1 \end{bmatrix}^{-1} \begin{pmatrix} \tilde{\mathbf{l}}' \\ 1 \end{pmatrix} = 0. \tag{11.12}$$

Finally, this is equivalent to

$$\mathbf{l}'^{\mathsf{T}}[\mathbf{m}'\mathbf{m}'^{\mathsf{T}} - k^2\Sigma'_{\mathbf{l}}]^{-1}\mathbf{l}' = 0. \tag{11.13}$$

This shows that the lines satisfying (11.11) form a line conic defined by the matrix $(\mathbf{m}'\mathbf{m}'^{\mathsf{T}} - k^2\Sigma'_{\mathbf{l}})^{-1}$. The corresponding point conic, which forms the envelope of the lines, is defined by the matrix $\mathbf{m}'\mathbf{m}'^{\mathsf{T}} - k^2\Sigma'_{\mathbf{l}}$. One may now transform back to the original coordinate system to determine the envelope of the lines in the original coordinate system. The transformed conic is

$$\mathtt{C} = \mathtt{U}^{\mathsf{T}}(\mathbf{m}'\mathbf{m}'^{\mathsf{T}} - k^2\Sigma'_{\mathbf{l}})\mathtt{U} = \mathbf{m}\mathbf{m}^{\mathsf{T}} - k^2\Sigma_{\mathbf{l}}. \tag{11.14}$$

Note that when $k = 0$, the conic $\mathtt{C}$ degenerates to $\mathbf{m}\mathbf{m}^{\mathsf{T}}$, which represents the set of points lying on the line $\mathbf{m}$. As k increases, the conic becomes a hyperbola the two branches of which lie on opposite sides of the line $\mathbf{m}$.

Suppose we want to choose k so that some fraction α of the epipolar lines lie inside the region bounded by this hyperbola. The value $k^2 = (\mathbf{l} - \mathbf{m})^{\mathsf{T}}\Sigma_{\mathbf{l}}^{+}(\mathbf{l} - \mathbf{m})$ of (11.11) follows a χ^2_n distribution, and the cumulative chi-squared distribution $F_n(k^2) = \int_0^{k^2} \chi^2_n(\xi)d\xi$ represents the probability that the value of a χ^2_n random variable is less than k^2 (the χ^2_n and F_n distributions are defined in section A2.2(*p*566)). Applying this to a random line $\mathbf{l}$, one sees that in order to ensure that a fraction α of lines lie within region bounded by the hyperbola defined by (11.14), one must choose k^2 such that $F_2(k^2) = \alpha$ ($n = 2$ since the covariance matrix $\Sigma_{\mathbf{l}}$ has rank 2). Thus, $k^2 = F_2^{-1}(\alpha)$, and for a value of $\alpha = 0.95$, for instance, one finds that $k^2 = 5.9915$. The corresponding hyperbola given by (11.14) is $\mathtt{C} = \mathbf{m}\mathbf{m}^{\mathsf{T}} - 5.9915\,\Sigma_{\mathbf{l}}$. To sum up this discussion:

Result 11.2. *If* $\mathbf{l}$ *is a random line obeying a Gaussian distribution with mean* $\bar{\mathbf{l}}$ *and covariance matrix* $\Sigma_{\mathbf{l}}$ *of rank 2, then the plane conic*

$$\mathtt{C} = \bar{\mathbf{l}}\,\bar{\mathbf{l}}^{\mathsf{T}} - k^2\Sigma_{\mathbf{l}} \tag{11.15}$$

represents an equal-likelihood contour bounding some fraction of all instances of $\mathbf{l}$. *If* $F_2(k^2)$ *represents the cumulative* χ^2_2 *distribution, and* k^2 *is chosen such that* $F_2^{-1}(k^2) = \alpha$, *then a fraction* α *of all lines lie within the region bounded by* $\mathtt{C}$. *In other words with probability* α *the lines lie within this region*.

In applying this formula, one must be aware that it represents only an approximation, since epipolar lines are not normally distributed. We have consistently made the assumption that the distributions may be correctly transformed using the Jacobian, that is an assumption of linearity. This assumption will be most reasonable for distributions with small variance, and close to the mean. Here, we are applying it to find the region in which as many as 95% of samples fall, namely almost the whole of the error

distribution. In this case, the assumption of a Gaussian distribution of errors is less tenable.

11.11.1 Verification of epipolar line covariance

We now present some examples of epipolar line envelopes, confirming and illustrating the theory developed above. Before doing this, however, a direct verification of the theory will be given, concerning the covariance matrix of epipolar lines. Since the 3×3 covariance matrix of a line is not easily understood quantitatively, we consider the variance of the direction of epipolar lines. Given a line $\mathbf{l} = (l_1, l_2, l_3)^\mathsf{T}$, the angle representing its direction is given by $\theta = \arctan(-l_1/l_2)$. Letting $\mathtt{J}$ equal the 1×3 Jacobian matrix of the mapping $\mathbf{l} \rightarrow \theta$, one finds the variance of the angle θ to be $\sigma_\theta^2 = \mathtt{J}\Sigma_\mathbf{l}\mathtt{J}^\mathsf{T}$. This result may be verified by simulation, as follows.

One considers a pair of images for which point correspondences have been identified. The fundamental matrix is computed from the point correspondences and the points are then corrected so as to correspond precisely under the epipolar mapping (as described in section 12.3). A set of n of these corrected correspondences are used to compute the covariance matrix of the fundamental matrix $\mathtt{F}$. Then, for a further set of "test" corrected points $\mathbf{x}_i$ in the first image, the mean and covariance of the corresponding epipolar line $\mathbf{l}'_i = \mathtt{F}\mathbf{x}_i$ are computed, and subsequently the mean and variance of the orientation direction of this line are computed. This gives the theoretical values of these quantities.

Next, Monte Carlo simulation is done, in which Gaussian noise is added to the coordinates of the points used to compute $\mathtt{F}$. Using the computed $\mathtt{F}$, the epipolar lines corresponding to each of the test points are computed, and subsequently their angle, and the deviation of the angle from the mean. This is done many times, and the standard deviation of angle is computed, and finally compared with the theoretical value. The results of this are shown in figure 11.8 for the statue image pair of figure 11.2-(*p*289).

Epipolar envelopes for statue image. The statue image pair of figure 11.2(*p*289) is interesting because of the large depth variation across the image. There are close points (on the statue) and distant points (on the building behind) in close proximity in the images. The fundamental matrix was computed from several points. A point in the first image (see figure 11.9) was selected and Monte Carlo simulation was used to compute several possible epipolar lines corresponding to a noise level of 0.5 pixels in each matched point coordinate. To test the theory, the mean and covariance of the epipolar line were next computed theoretically. The 95% envelope of the epipolar lines was computed and drawn in the second image. The results are shown in figure 11.10 for different numbers of points used to compute $\mathtt{F}$. The 95% envelope for $n = 15$ corresponds closely to the simulated envelope of the lines.

The results shown in figure 11.10 show the practical importance of computing the epipolar envelopes in point matching. Thus, suppose one is attempting to find a match for the foreground point in figure 11.9. If the epipolar line is computed from just 10 point matches, then epipolar search is unlikely to succeed, given the width of the

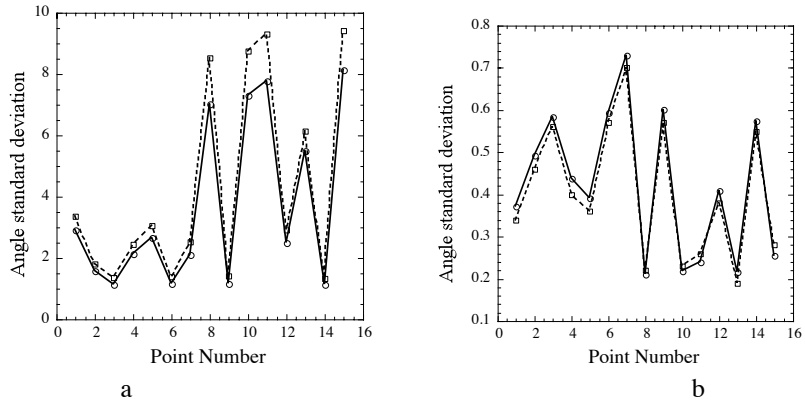


Fig. 11.8. *Comparison of theoretical and Monte Carlo simulated values of orientation angle of epipolar lines for 15 test points form the statue image pair of figure 11.2(*p289*). The horizontal axis represents the point number (1 to 15) and the vertical axis the standard deviation of angle. (a) the results when the epipolar structure (fundamental matrix) is computed from 15 points. (b) the results when 50 point matches are used.* **Note :** *the horizontal axis of these graphs represent discrete points numbered 1 to 15. The graphs are shown as a continuous curve only for visual clarity.*

envelope. Even for $n = 15$, the width of the envelope at the level of the correct match is several tens of pixels. For $n = 25$, the situation is more favourable. Note that this instability is inherent in the problem, and not the result of any specific algorithm for computing F.

An interesting point concerns the location of the narrowest point of the envelope. In this case, it appears to be close to the correct match position for the background point in figure 11.9. The match for the foreground point (leg of statue) lies far from the narrowest point of the envelope. Though the precise location of the narrow point of the envelope is not fully understood, it appears that in this case, most points used to the computation of F are on the background building. This biases towards the supposition that other matched points lie close to the plane of the building. The match for a point at significantly different depth is less precisely known.

Matching points close to the epipole – the corridor scene. In the case where the points to be matched are close to the epipole, then the determination of the epipolar line is more unstable, since any uncertainty in the position of the epipole results in uncertainty in the slope of the epipolar line. In addition, as one approaches this unstable position, the linear approximations implicit in the derivation of (11.14) become less tenable. In particular, the distribution of the epipolar lines deviates from a normal distribution.

11.12 Image rectification

This section gives a method for image rectification, the process of resampling pairs of stereo images taken from widely differing viewpoints in order to produce a pair of "matched epipolar projections". These are projections in which the epipolar lines run parallel with the x-axis and match up between views, and consequently disparities between the images are in the x-direction only, i.e. there is no y disparity.

a b

Fig. 11.9. *(a) The point in the first image used to compute the epipolar envelopes in the second images. Note the ambiguity of which point is to be found in the second image. The marked point may represent the point on the statue's leg (foreground) or the point on the building behind the statue (background). In the second image, these two points are quite separate, and the epipolar line must pass through them both. (b) Computed corresponding epipolar lines computed from $n = 15$ point matches. The different lines correspond to different instances of injected noise in the matched points. Gaussian noise of 0.5 pixels in each coordinate was added to the ideal matched point positions before computing the epipolar line corresponding to the selected point. The ML estimator (Gold Standard algorithm) was used to compute* `F`*. This experiment shows the basic instability of the computation of the epipolar lines from small numbers of points. To find the point matching the selected point in the image at left, one needs to search over the regions covered by all these epipolar lines.*

The method is based on the fundamental matrix. A pair of 2D projective transformations are applied to the two images in order to match the epipolar lines. It is shown that the two transformations may be chosen in such a way that matching points have approximately the same x-coordinate as well. In this way, the two images, if overlaid on top of each other, will correspond as far as possible, and any disparities will be parallel to the x-axis. Since the application of arbitrary 2D projective transformations may distort the image substantially, the method for finding the pair of transformations subjects the images to a minimal distortion.

In effect, transforming the two images by the appropriate projective transformations reduces the problem to the epipolar geometry produced by a pair of identical cameras placed side by side with their principal axes parallel. Many stereo matching algorithms described in previous literature have assumed this geometry. After this rectification the search for matching points is vastly simplified by the simple epipolar structure and by the near-correspondence of the two images. It may be used as a preliminary step to comprehensive image matching.

11.12.1 Mapping the epipole to infinity

In this section we will discuss the question of finding a projective transformation `H` of an image that maps the epipole to a point at infinity. In fact, if epipolar lines are to be transformed to lines parallel with the x-axis, then the epipole should be mapped to

Fig. 11.10. *The 95% envelopes of epipolar lines are shown for a noise level of* 0.5 *pixels, with* F *being computed from* $n = 10, 15, 25$ *and* 50 *points. In each case, Monte Carlo simulated results agreed closely with these results (though not shown here). For the case* $n = 15$*, compare with figure 11.9. Note that for* $n = 10$*, the epipolar envelope is very wide (> 90 degrees), showing that one can have very little confidence in an epipolar line computed from 10 points in this case. For* $n = 15$*, the envelope is still quite wide. For* $n = 25$ *and* $n = 50$*, the epipolar line is known with quite good precision. Of course, the precise shape of the envelope depends strongly on just what matched points are used to compute the epipolar structure.*

the particular infinite point $(1, 0, 0)^\mathsf{T}$. This leaves many degrees of freedom (in fact four) open for H, and if an inappropriate H is chosen, severe projective distortion of the image can take place. In order that the resampled image should look somewhat like the original image, we may put closer restrictions on the choice of H.

One condition that leads to good results is to insist that the transformation H should act as far as possible as a rigid transformation in the neighbourhood of a given selected point $\mathbf{x}_0$ of the image. By this is meant that to first-order the neighbourhood of $\mathbf{x}_0$ may undergo rotation and translation only, and hence will look the same in the original and resampled images. An appropriate choice of point $\mathbf{x}_0$ may be the centre of the image. For instance, this would be a good choice in the context of aerial photography if the view is known not to be excessively oblique.

For the present, suppose $\mathbf{x}_0$ is the origin and the epipole $\mathbf{e} = (f, 0, 1)^\mathsf{T}$ lies on the

x-axis. Now consider the following transformation

$$\mathtt{G} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1/f & 0 & 1 \end{bmatrix}. \tag{11.16}$$

This transformation takes the epipole $(f, 0, 1)^\mathsf{T}$ to the point at infinity $(f, 0, 0)^\mathsf{T}$ as required. A point $(x, y, 1)^\mathsf{T}$ is mapped by $\mathtt{G}$ to the point $(\hat{x}, \hat{y}, 1)^\mathsf{T} = (x, y, 1 - x/f)^\mathsf{T}$. If $|x/f| < 1$ then we may write

$$(\hat{x}, \hat{y}, 1)^\mathsf{T} = (x, y, 1 - x/f)^\mathsf{T} = (x(1 + x/f + \ldots), y(1 + x/f + \ldots), 1)^\mathsf{T}.$$

The Jacobian is

$$\frac{\partial(\hat{x}, \hat{y})}{\partial(x, y)} = \begin{bmatrix} 1 + 2x/f & 0 \\ y/f & 1 + x/f \end{bmatrix}$$

plus higher order terms in x and y. Now if $x = y = 0$ then this is the identity map. In other words, $\mathtt{G}$ is approximated (to first-order) at the origin by the identity mapping.

For an arbitrarily placed point of interest $\mathbf{x}_0$ and epipole $\mathbf{e}$, the required mapping $\mathtt{H}$ is a product $\mathtt{H} = \mathtt{GRT}$ where $\mathtt{T}$ is a translation taking the point $\mathbf{x}_0$ to the origin, $\mathtt{R}$ is a rotation about the origin taking the epipole $\mathbf{e}'$ to a point $(f, 0, 1)^\mathsf{T}$ on the x-axis, and $\mathtt{G}$ is the mapping just considered taking $(f, 0, 1)^\mathsf{T}$ to infinity. The composite mapping is to first-order a rigid transformation in the neighbourhood of $\mathbf{x}_0$.

11.12.2 Matching transformations

In the previous section it was shown how the epipole in one image may be mapped to infinity. Next, it will be seen how a map may be applied to the other image to match up the epipolar lines. We consider two images J and J'. The intention is to resample these two images according to transformations $\mathtt{H}$ to be applied to J and $\mathtt{H}'$ to be applied to J'. The resampling is to be done in such a way that an epipolar line in J is matched with its corresponding epipolar line in J'. More specifically, if $\mathbf{l}$ and $\mathbf{l}'$ are any pair of corresponding epipolar lines in the two images, then $\mathtt{H}^{-\mathsf{T}}\mathbf{l} = \mathtt{H}'^{-\mathsf{T}}\mathbf{l}'$. (Recall that $\mathtt{H}^{-\mathsf{T}}$ is the line map corresponding to the point map $\mathtt{H}$.) Any pair of transformations satisfying this condition will be called a *matched pair* of transformations.

Our strategy in choosing a matched pair of transformations is to choose $\mathtt{H}'$ first to be some transformation that sends the epipole $\mathbf{e}'$ to infinity as described in the previous section. We then seek a matching transformation $\mathtt{H}$ chosen so as to minimize the sum-of-squared distances

$$\sum_i d(\mathtt{H}\mathbf{x}_i, \mathtt{H}'\mathbf{x}'_i)^2. \tag{11.17}$$

The first question to be determined is how to find a transformation matching $\mathtt{H}'$. That question is answered in the following result.

Result 11.3. *Let J and J' be images with fundamental matrix $\mathtt{F} = [\mathbf{e}']_\times\mathtt{M}$, and let $\mathtt{H}'$ be a projective transformation of J'. A projective transformation $\mathtt{H}$ of J matches $\mathtt{H}'$ if and*

only if $\mathtt{H}$ *is of the form*

$$\mathtt{H} = (\mathtt{I} + \mathtt{H}'\mathbf{e}'\mathbf{a}^\mathsf{T})\mathtt{H}'\mathtt{M} \tag{11.18}$$

for some vector $\mathbf{a}$.

Proof. If $\mathbf{x}$ is a point in J, then $\mathbf{e}\times\mathbf{x}$ is the epipolar line in the first image, and $\mathtt{F}\mathbf{x}$ is the epipolar line in the second image. Transformations $\mathtt{H}$ and $\mathtt{H}'$ are a matching pair if and only if $\mathtt{H}^{-\mathsf{T}}(\mathbf{e}\times\mathbf{x}) = \mathtt{H}'^{-\mathsf{T}}\mathtt{F}\mathbf{x}$. Since this must hold for all $\mathbf{x}$ we may write equivalently $\mathtt{H}^{-\mathsf{T}}[\mathbf{e}]_\times = \mathtt{H}'^{-\mathsf{T}}\mathtt{F} = \mathtt{H}'^{-\mathsf{T}}[\mathbf{e}']_\times\mathtt{M}$ or, applying result A4.3(*p*582),

$$[\mathtt{H}\mathbf{e}]_\times\mathtt{H} = [\mathtt{H}'\mathbf{e}']_\times\mathtt{H}'\mathtt{M}. \tag{11.19}$$

In view of lemma 9.11(*p*255), this implies $\mathtt{H} = (\mathtt{I} + \mathtt{H}'\mathbf{e}'\mathbf{a}^\mathsf{T})\mathtt{H}'\mathtt{M}$ as required.
To prove the converse, if (11.18) holds, then

$$\begin{aligned}\mathtt{H}\mathbf{e} &= (\mathtt{I} + \mathtt{H}'\mathbf{e}'\mathbf{a}^\mathsf{T})\mathtt{H}'\mathtt{M}\mathbf{e} = (\mathtt{I} + \mathtt{H}'\mathbf{e}'\mathbf{a}^\mathsf{T})\mathtt{H}'\mathbf{e}' \\ &= (1 + \mathbf{a}^\mathsf{T}\mathtt{H}'\mathbf{e}')\mathtt{H}'\mathbf{e}' = \mathtt{H}'\mathbf{e}'.\end{aligned}$$

This, along with (11.18), is sufficient for (11.19) to hold, and so $\mathtt{H}$ and $\mathtt{H}'$ are matching transformations. □

We are particularly interested in the case when $\mathtt{H}'$ is a transformation taking the epipole $\mathbf{e}'$ to a point at infinity $(1,0,0)^\mathsf{T}$. In this case, $\mathtt{I} + \mathtt{H}'\mathbf{e}'\mathbf{a}^\mathsf{T} = \mathtt{I} + (1,0,0)^\mathsf{T}\mathbf{a}^\mathsf{T}$ is of the form

$$\mathtt{H}_\mathrm{A} = \begin{bmatrix} a & b & c \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{11.20}$$

which represents an affine transformation. Thus, a special case of result 11.3 is

Corollary 11.4. *Let* J *and* J' *be images with fundamental matrix* $\mathtt{F} = [\mathbf{e}']_\times\mathtt{M}$*, and let* $\mathtt{H}'$ *be a projective transformation of* J' *mapping the epipole* $\mathbf{e}'$ *to the infinite point* $(1,0,0)^\mathsf{T}$*. A transformation* $\mathtt{H}$ *of* J *matches* $\mathtt{H}'$ *if and only if* $\mathtt{H}$ *is of the form* $\mathtt{H} = \mathtt{H}_\mathrm{A}\mathtt{H}_0$*, where* $\mathtt{H}_0 = \mathtt{H}'\mathtt{M}$ *and* $\mathtt{H}_\mathrm{A}$ *is an affine transformation of the form (11.20).*

Given $\mathtt{H}'$ mapping the epipole to infinity, we may use this corollary to make the choice of a matching transformation $\mathtt{H}$ to minimize the disparity. Writing $\hat{\mathbf{x}}'_i = \mathtt{H}'\mathbf{x}'_i$ and $\hat{\mathbf{x}}_i = \mathtt{H}_0\mathbf{x}_i$, the minimization problem (11.17) is to find $\mathtt{H}_\mathrm{A}$ of the form (11.20) such that

$$\sum_i d(\mathtt{H}_\mathrm{A}\hat{\mathbf{x}}_i, \hat{\mathbf{x}}'_i)^2 \tag{11.21}$$

is minimized.

In particular, let $\hat{\mathbf{x}}_i = (\hat{x}_i, \hat{y}_i, 1)^\mathsf{T}$, and let $\hat{\mathbf{x}}'_i = (\hat{x}'_i, \hat{y}'_i, 1)^\mathsf{T}$. Since $\mathtt{H}'$ and $\mathtt{M}$ are known, these vectors may be computed from the matched points $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$. Then the quantity to be minimized (11.21) may be written as

$$\sum_i (a\hat{x}_i + b\hat{y}_i + c - \hat{x}'_i)^2 + (\hat{y}_i - \hat{y}'_i)^2.$$

Since $(\hat{y}_i - \hat{y}'_i)^2$ is a constant, this is equivalent to minimizing

$$\sum_i (a\hat{x}_i + b\hat{y}_i + c - \hat{x}'_i)^2.$$

This is a simple linear least-squares parameter minimization problem, and is easily solved using linear techniques (see section A5.1(*p*588)) to find a, b and c. Then H_{A} is computed from (11.20) and $\mathtt{H}$ from (11.18). Note that a linear solution is possible because $\mathtt{H}_{\mathrm{A}}$ is an affine transformation. If it were simply a projective transformation, this would not be a linear problem.

11.12.3 Algorithm outline

The resampling algorithm will now be summarized. The input is a pair of images containing a common overlap region. The output is a pair of images resampled so that the epipolar lines in the two images are horizontal (parallel with the x-axis), and such that corresponding points in the two images are as close to each other as possible. Any remaining disparity between matching points will be along the the horizontal epipolar lines. A top-level outline of the algorithm is as follows.

(i) Identify a seed set of image-to-image matches $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ between the two images. Seven points at least are needed, though more are preferable. It is possible to find such matches by automatic means.

(ii) Compute the fundamental matrix $\mathtt{F}$ and find the epipoles $\mathbf{e}$ and $\mathbf{e}'$ in the two images.

(iii) Select a projective transformation $\mathtt{H}'$ that maps the epipole $\mathbf{e}'$ to the point at infinity, $(1, 0, 0)^\mathsf{T}$. The method of section 11.12.1 gives good results.

(iv) Find the matching projective transformation $\mathtt{H}$ that minimizes the least-squares distance

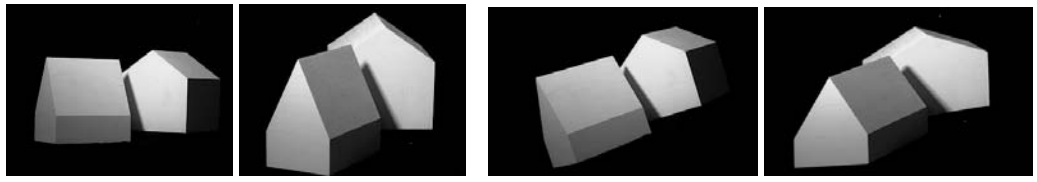
$$\sum_i d(\mathtt{H}\mathbf{x}_i, \mathtt{H}'\mathbf{x}'_i). \tag{11.22}$$

The method used is a linear method described in section 11.12.2.

(v) Resample the first image according to the projective transformation $\mathtt{H}$ and the second image according to the projective transformation $\mathtt{H}'$.

Example 11.5. Model house images

Figure 11.11(a) shows a pair of images of some wooden block houses. Edges and vertices in these two images were extracted automatically and a small number of common vertices were matched by hand. The two images were then resampled according to the methods described here. The results are shown in figure 11.11(b). In this case, because of the wide difference in viewpoint, and the three-dimensional shape of the objects, the two images even after resampling look quite different. However, it is the case that any point in the first image will now match a point in the second image with the same y-coordinate. Therefore, in order to find further point matches between the images only a 1-dimensional search is required. △



a b

Fig. 11.11. **Image rectification.** *(a) A pair of images of a house. (b) Resampled images computed from (a) using a projective transformation computed from* F*. Note, corresponding points in (b) match horizontally.*

a b

Fig. 11.12. **Image rectification using affinities.** *(a) A pair of original images and (b) a detail of the images rectified using affine transformations. The average y-disparity after rectification is of the order of 3 pixels in a 512×512 image. (For correctly rectified images the y-disparity should be zero.)*

11.12.4 Affine rectification

The theory discussed in this section can equally be applied to affine resampling. If the two cameras can be well approximated by affine cameras, then one can rectify the images using just affine transformations. To do this, one uses the affine fundamental matrix (see section 14.2(*p*345)) instead of the general fundamental matrix. The above method with only minor variations can then be applied to compute a pair of matching affine transformations. Figure 11.12 shows a pair of images rectified using affine transformations.

11.13 Closure

11.13.1 The literature

The basic idea behind the computation of the fundamental matrix is given in [LonguetHiggins-81], which is well worth reading. It addresses the case of calibrated matrices only, but the principles apply to the uncalibrated case as well. A good reference for the uncalibrated case is [Zhang-98] which considers most of the best methods. In addition, that paper considers the uncertainty envelopes of epipolar lines, following earlier work by Csurka *et al.* [Csurka-97]. A more detailed study of the 8-point algorithm in the uncalibrated case is given in [Hartley-97c]. Weng *et al.* [Weng-89] used Sampson approximation for the fundamental matrix cost function. The SVD method of coercing the estimated F to have rank 2 was suggested by Tsai & Huang [Tsai-84].

There is a wealth of literature on conic fitting – minimizing algebraic distance

[Bookstein-79]; approximating geometric distance [Sampson-82, Pratt-87, Taubin-91]; optimal fitting [Kanatani-94]; and fitting special forms [Fitzgibbon-99].

11.13.2 Notes and exercises

(i) Six point correspondences constrain $\mathbf{e}$ and $\mathbf{e}'$ to a plane cubic in each image ([Faugeras-93], page 298). The cubic also passes through the six points in each image. A sketch derivation of these results follows. Given six correspondences, the null-space of $\mathtt{A}$ in (11.3–*p*279) will be 3-dimensional. Then the solution is $\mathtt{F} = \alpha_1\mathtt{F}_1 + \alpha_2\mathtt{F}_2 + \alpha_3\mathtt{F}_3$, where $\mathtt{F}_i$ denotes the matrices corresponding to the vectors spanning the null-space. The epipole satisfies $\mathtt{F}\mathbf{e} = \mathbf{0}$, so that $[(\mathtt{F}_1\mathbf{e}), (\mathtt{F}_2\mathbf{e}), (\mathtt{F}_3\mathbf{e})](\alpha_1, \alpha_2, \alpha_3)^\mathsf{T} = \mathbf{0}$. Since this equation has a solution it follows that $\det[(\mathtt{F}_1\mathbf{e}), (\mathtt{F}_2\mathbf{e}), (\mathtt{F}_3\mathbf{e})] = 0$ which is a cubic in $\mathbf{e}$.

(ii) Show that the image correspondence of four coplanar points and a quadric outline determines the fundamental matrix up to a two-fold ambiguity (Hint, see algorithm 13.2(*p*336)).

(iii) Show that the corresponding images of a (plane) conic are equivalent to two constraints on $\mathtt{F}$. See [Kahl-98b] for details.

(iv) Suppose that a stereo pair of images is acquired by a camera translating forward along its principal axis. Can the geometry of image rectification described in section 11.12 be applied in this case? See [Pollefeys-99a] for an alternative rectification geometry.