14

Affine Epipolar Geometry

This chapter recapitulates the developments and objectives of the previous chapters on two-view geometry, but here with affine cameras replacing projective cameras. The affine camera is an extremely usable and well conditioned approximation in many practical situations. Its great advantage is that, because of its linearity, many of the optimal algorithms can be implemented by linear algebra (matrix inverses, SVD etc.), whereas in the projective case solutions either involve high order polynomials (such as for triangulation) or are only possible by numerical minimization (such as in the Gold Standard estimation of F).

We first describe properties of the epipolar geometry of two affine cameras, and its optimal computation from point correspondences. This is followed by triangulation, and *affine* reconstruction. Finally the ambiguities in reconstruction that result from parallel projection are sketched, and the non-ambiguous motion parameters are computed from the epipolar geometry.

14.1 Affine epipolar geometry

In many respects the epipolar geometry of two affine cameras is identical to that of two perspective cameras, for example a point in one view defines an epipolar line in the other view, and the pencil of such epipolar lines intersect at the epipole. The difference is that because the cameras are affine their centres are at infinity, and there is parallel projection from scene to image. This leads to certain simplifications in the affine epipolar geometry:

Epipolar lines. Consider two points, $\mathbf{x}_1$, $\mathbf{x}_2$, in the first view. These points back-project to rays which are *parallel* in 3-space, since all projection rays are parallel. In the second view an epipolar line is the image of a back-projected ray. The images of these two rays in the second view are also parallel since an affine camera maps parallel scene lines to parallel images lines. Consequently, all epipolar lines are parallel, as are the epipolar planes.

The epipoles. Since epipolar lines intersect in the epipole, and all epipolar lines are parallel, it follows that the epipole is at infinity.

Downloaded from https://www.cambridge.org/core. University of Illinois at Urbana - Champaign Library, on 22 Sep 2018 at 15:36:38, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811685.019

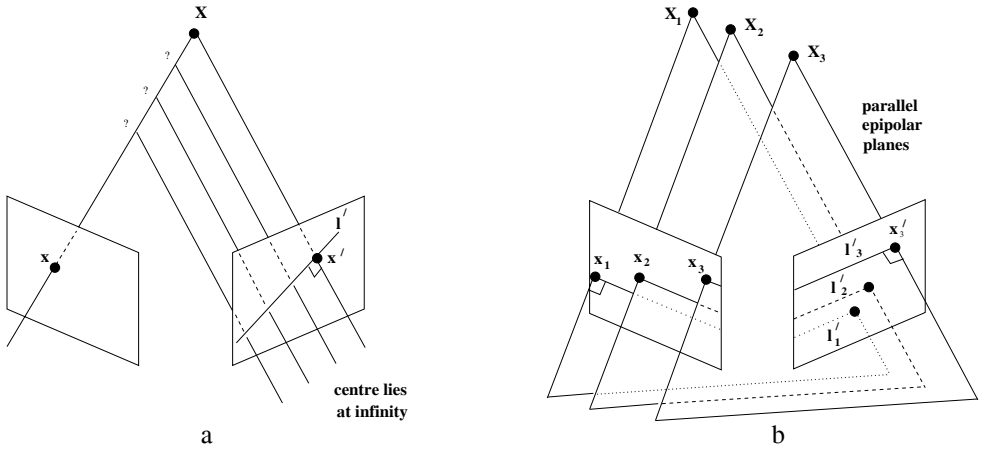


Fig. 14.1. **Affine epipolar geometry.** *(a) Correspondence geometry: Projection rays are parallel and intersect at infinity. A point* $\mathbf{x}$ *back-projects to a ray in 3-space defined by the first camera centre (at infinity) and* $\mathbf{x}$*. This ray is imaged as a line* $\mathbf{l}'$ *in the second view. The 3-space point* $\mathbf{X}$ *which projects to* $\mathbf{x}$ *lies on this ray, so the image of* $\mathbf{X}$ *in the second view lies on* $\mathbf{l}'$*. (b) Epipolar lines and planes are parallel.*

These points are illustrated schematically in figure 14.1, and examples on images are shown in figure 14.2.

14.2 The affine fundamental matrix

The affine epipolar geometry is represented algebraically by a matrix termed the *affine fundamental matrix*, $\mathtt{F}_{\mathrm{A}}$. It will be seen in the following that:

Result 14.1. *The fundamental matrix resulting from two cameras with the affine form (i.e. the third row is* $(0, 0, 0, 1)$*) has the form*

$$\mathtt{F}_{\mathrm{A}} = \begin{bmatrix} 0 & 0 & * \\ 0 & 0 & * \\ * & * & * \end{bmatrix}$$

where $*$ *indicates a non-zero entry.*

It will be convenient to write the five non-zero entries as

$$\mathtt{F}_{\mathrm{A}} = \begin{bmatrix} 0 & 0 & a \\ 0 & 0 & b \\ c & d & e \end{bmatrix}. \tag{14.1}$$

Note that in general $\mathtt{F}_{\mathrm{A}}$ has rank 2.

14.2.1 Derivation

Geometric derivation. This derivation is the analogue of that given in section 9.2.1-(*p*242) for a pair of projective cameras. The map from a point in one image to the corresponding epipolar line in the other image is decomposed into two steps, as illustrated in figure 14.5 on page 352:

Downloaded from https://www.cambridge.org/core. University of Illinois at Urbana - Champaign Library, on 22 Sep 2018 at 15:36:38, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811685.019

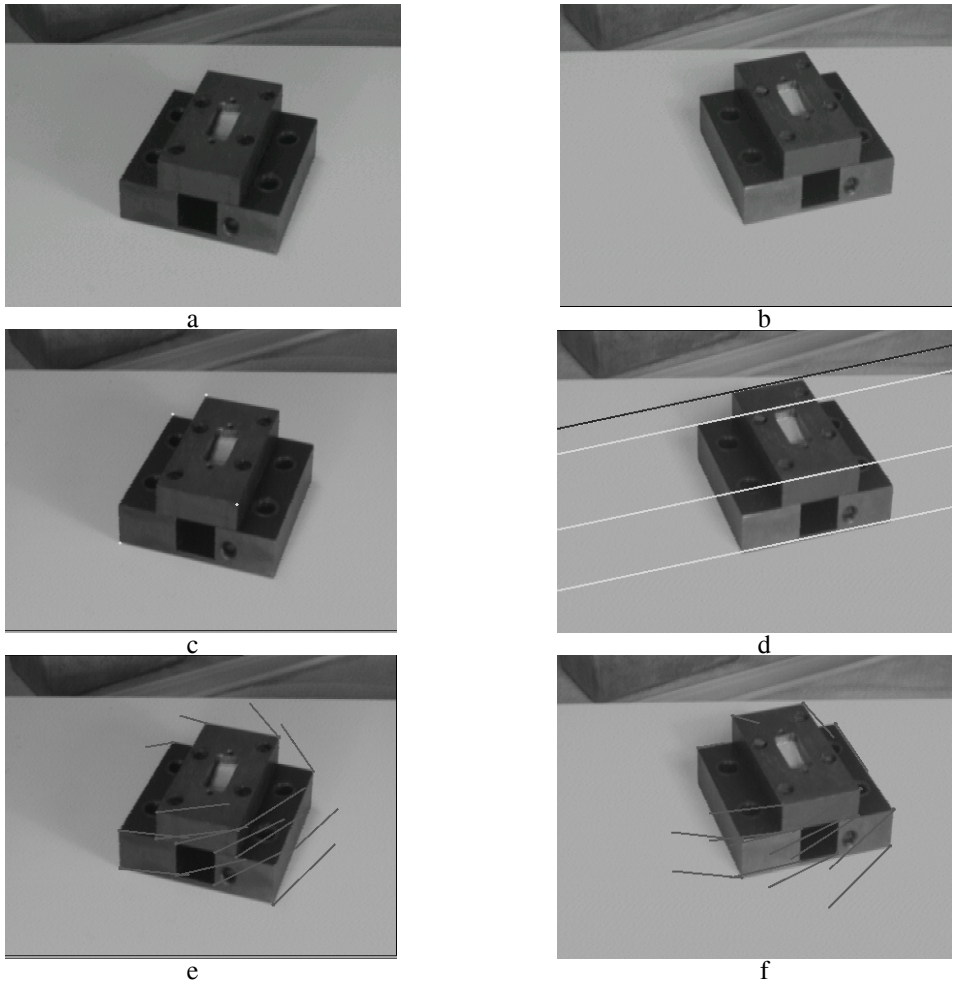


Fig. 14.2. **Affine epipolar lines.** *(a), (b) Two views of a hole punch acquired under affine imaging conditions. For the points marked in (c) the epipolar lines are superimposed on (d). Note that corresponding points lie on their epipolar lines, and that all epipolar lines are parallel. The epipolar geometry is computed from point correspondences using algorithm 14.1. (e) and (f) show the "flow" for selected points in the image (the lines link a point in one image to the point's position in the other image). This demonstrates that even though the epipolar lines are parallel the movement of imaged points between the views contains both rotational and translational components.*

(i) **Point transfer via a plane π.** Since both cameras are affine, points are mapped between an image and a scene plane by parallel projection, so the map between π and the images is a planar affine transformation; the composition of the affine transformations between the first view and π, and π and the second view, is also an affine transformation, i.e. $\mathbf{x}' = \mathtt{H}_{\mathrm{A}}\mathbf{x}$.

(ii) **Constructing the epipolar line.** The epipolar line is obtained as the line through $\mathbf{x}'$ and the epipole $\mathbf{e}'$, i.e. $\mathbf{l}' = \mathbf{e}' \times \mathtt{H}_{\mathrm{A}}\mathbf{x} = \mathtt{F}_{\mathrm{A}}\mathbf{x}$, so that $\mathtt{F}_{\mathrm{A}} = [\mathbf{e}']_{\times}\mathtt{H}_{\mathrm{A}}$.

Downloaded from https://www.cambridge.org/core. University of Illinois at Urbana - Champaign Library, on 22 Sep 2018 at 15:36:38, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811685.019

We now take note of the special forms of the affine matrix $\mathtt{H}_{\mathrm{A}}$, and the skew matrix $[\mathbf{e}']_\times$ when $\mathbf{e}'$ is at infinity, and so has a zero last element:

$$\mathtt{F}_{\mathrm{A}} = [\mathbf{e}']_\times \mathtt{H}_{\mathrm{A}} = \begin{bmatrix} 0 & 0 & * \\ 0 & 0 & * \\ * & * & 0 \end{bmatrix} \begin{bmatrix} * & * & * \\ * & * & * \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & * \\ 0 & 0 & * \\ * & * & * \end{bmatrix} \tag{14.2}$$

where $*$ indicates a non-zero entry. This derives the affine form of $\mathtt{F}$ using only the geometric properties that the camera centres are on the plane at infinity.

Algebraic derivation. In the case that the cameras are both affine, the affine form of the fundamental matrix is obtained directly from the expression (9.1–*p*244) for $\mathtt{F}$ in terms of the pseudo-inverse, namely $\mathtt{F} = [\mathbf{e}']_\times \mathtt{P}'\mathtt{P}^+$, where $\mathbf{e}' = \mathtt{P}'\mathbf{C}$, with $\mathbf{C}$ the camera centre which is the null-vector of $\mathtt{P}$. Details are left as an exercise. An elegant derivation of $\mathtt{F}_{\mathrm{A}}$ in terms of determinants formed from rows of the affine camera matrices is given in section 17.1.2(*p*413).

14.2.2 Properties

The affine fundamental matrix is a homogeneous matrix with five non-zero elements, it thus has 4 degrees of freedom. These are accounted as: one for each of the two epipoles (the epipoles lie on $\mathbf{l}_\infty$, so only their direction need be specified); and two for the 1D affine transformation mapping the pencil of epipolar lines from one view to the other.

The geometric entities (epipoles etc.) are encoded in $\mathtt{F}_{\mathrm{A}}$ in the same manner as their encoding in $\mathtt{F}$. However, often the expressions are far simpler and so can be given explicitly.

The epipoles. The epipole in the first view is the right null-vector of $\mathtt{F}_{\mathrm{A}}$, i.e. $\mathtt{F}_{\mathrm{A}}\mathbf{e} = \mathbf{0}$. This determines $\mathbf{e} = (-d, c, 0)^\mathsf{T}$, which is a point (direction) on $\mathbf{l}_\infty$. Since all epipolar lines intersect the epipole this shows that all epipolar lines are parallel.

Epipolar lines. The epipolar line in the second view corresponding to $\mathbf{x}$ in the first is $\mathbf{l}' = \mathtt{F}_{\mathrm{A}}\mathbf{x} = (a, b, cx + dy + e)^\mathsf{T}$. Again it is evident that all epipolar lines are parallel since the line orientation, (a, b), is independent of (x, y).
These properties, and others, are summarized in table 14.1.

14.3 Estimating $\mathtt{F}_{\mathrm{A}}$ from image point correspondences

The fundamental matrix is defined by the equation $\mathbf{x}'^\mathsf{T}\mathtt{F}_{\mathrm{A}}\mathbf{x} = 0$ for any pair of matching points $\mathbf{x} \leftrightarrow \mathbf{x}'$ in two images. Given sufficiently many point matches $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$, this equation can be used to compute the unknown matrix $\mathtt{F}_{\mathrm{A}}$. In particular, writing $\mathbf{x}_i = (x_i, y_i, 1)^\mathsf{T}$ and $\mathbf{x}'_i = (x'_i, y'_i, 1)^\mathsf{T}$ each point match gives rise to one linear equation

$$ax'_i + by'_i + cx_i + dy_i + e = 0 \tag{14.3}$$

in the unknown entries $\{a, b, c, d, e\}$ of $\mathtt{F}_{\mathrm{A}}$.

Downloaded from https://www.cambridge.org/core. University of Illinois at Urbana - Champaign Library, on 22 Sep 2018 at 15:36:38, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811685.019

- $\mathtt{F}_{\mathrm{A}}$ is a rank 2 homogeneous matrix with 4 degrees of freedom. It has the form

$$\mathtt{F}_{\mathrm{A}} = \begin{bmatrix} 0 & 0 & a \\ 0 & 0 & b \\ c & d & e \end{bmatrix}.$$

- **Point correspondence**: If $\mathbf{x}$ and $\mathbf{x}'$ are corresponding image points under an affine camera, then $\mathbf{x}'^{\mathsf{T}}\mathtt{F}_{\mathrm{A}}\mathbf{x} = 0$. For finite points

$$ax' + by' + cx + dy + e = 0.$$

- **Epipolar lines**:
  - $\mathbf{l}' = \mathtt{F}_{\mathrm{A}}\mathbf{x} = (a, b, cx + dy + e)^{\mathsf{T}}$ is the epipolar line corresponding to $\mathbf{x}$.
  - $\mathbf{l} = \mathtt{F}_{\mathrm{A}}{}^{\mathsf{T}}\mathbf{x}' = (c, d, ax + by + e)^{\mathsf{T}}$ is the epipolar line corresponding to $\mathbf{x}'$.
- **Epipoles**:
  - From $\mathtt{F}_{\mathrm{A}}\mathbf{e} = \mathbf{0}$, $\mathbf{e} = (-d, c, 0)^{\mathsf{T}}$.
  - From $\mathtt{F}_{\mathrm{A}}{}^{\mathsf{T}}\mathbf{e}' = \mathbf{0}$, $\mathbf{e}' = (-b, a, 0)^{\mathsf{T}}$.
- **Computation from camera matrices** $\mathtt{P}_{\mathrm{A}}, \mathtt{P}'_{\mathrm{A}}$:
  - General cameras,
    $\mathtt{F}_{\mathrm{A}} = [\mathbf{e}']_{\times}\mathtt{P}'_{\mathrm{A}}\mathtt{P}^{+}_{\mathrm{A}}$, where $\mathtt{P}^{+}_{\mathrm{A}}$ is the pseudo-inverse of $\mathtt{P}_{\mathrm{A}}$, and $\mathbf{e}'$ is the epipole defined by $\mathbf{e}' = \mathtt{P}'_{\mathrm{A}}\mathbf{C}$, where $\mathbf{C}$ is the centre of the first camera.
  - Canonical cameras,

$$\mathtt{P}_{\mathrm{A}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \mathtt{P}'_{\mathrm{A}} = \left[\begin{array}{ccc|c} & \mathtt{M}_{2\times 3} & & \mathbf{t} \\ 0 & 0 & 0 & 1 \end{array}\right]$$

$$\begin{aligned} a &= m_{23}, \quad b = -m_{13}, \quad c = m_{13}m_{21} - m_{11}m_{23} \\ d &= m_{13}m_{22} - m_{12}m_{23}, \quad e = m_{13}t_2 - m_{23}t_1 \end{aligned}$$

Table 14.1. *Summary of affine fundamental matrix properties.*

14.3.1 The linear algorithm

In the usual manner a solution for $\mathtt{F}_{\mathrm{A}}$ may be obtained by rewriting (14.3) as

$$(x'_i, y'_i, x_i, y_i, 1)\,\mathbf{f} = 0$$

with $\mathbf{f} = (a, b, c, d, e)^{\mathsf{T}}$. From a set of n point matches, we obtain a set of linear equations of the form $\mathtt{A}\mathbf{f} = \mathbf{0}$, where $\mathtt{A}$ is a $n \times 5$ matrix:

$$\begin{bmatrix} x'_1 & y'_1 & x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x'_n & y'_n & x_n & y_n & 1 \end{bmatrix} \mathbf{f} = \mathbf{0}.$$

A minimal solution is obtained for $n = 4$ point correspondences as the right null-space of the 4×5 matrix $\mathtt{A}$. Thus $\mathtt{F}_{\mathrm{A}}$ can be computed uniquely from only 4 point

Downloaded from https://www.cambridge.org/core. University of Illinois at Urbana - Champaign Library, on 22 Sep 2018 at 15:36:38, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811685.019

correspondences, provided the 3-space points are in general position. The conditions for general position are described in section 14.3.3 below.

If there are more than 4 correspondences, and the data is not exact, then the rank of $\mathtt{A}$ may be greater than 4. In this case, one may find a least-squares solution, subject to $\|\mathbf{f}\| = 1$, in essentially the same manner as that of section 4.1.1(*p*90), as the singular vector corresponding to the smallest singular value of $\mathtt{A}$. Refer to algorithm 4.2(*p*109) for details. This linear solution is the equivalent of the 8-point algorithm 11.1(*p*282) for the computation of a general fundamental matrix. We do not recommend this approach for estimating $\mathtt{F}_\mathrm{A}$ because the Gold Standard algorithm described below may be implemented with equal computational ease, and in general will have superior performance.

The singularity constraint. The form (14.1) of $\mathtt{F}_\mathrm{A}$ ensures that the matrix has rank no greater than 2. Consequently, if $\mathtt{F}_\mathrm{A}$ is estimated by the linear method above it is not necessary to subsequently impose a singularity constraint. This is a considerable advantage over the estimation of a general $\mathtt{F}$ by the linear 8-point algorithm, where the estimated matrix is not guaranteed to have rank 2, and thus must be subsequently corrected.

Geometric interpretation. As has been seen at several points in this book, computing a two-view relation from point correspondences is equivalent to fitting a surface (variety) to points x, y, x', y' in $\mathbb{R}^4$. In the case of the equation $\mathbf{x}'^\mathsf{T}\mathtt{F}_\mathrm{A}\mathbf{x} = 0$ the relation $ax'_i + by'_i + cx_i + dy_i + e = 0$ is *linear* in the coordinates, and the variety $\mathcal{V}_{\mathtt{F}_\mathrm{A}}$ defined by the affine fundamental matrix is a hyperplane.

This results in two simplifications: first, finding the best estimate of $\mathtt{F}_\mathrm{A}$ may be formulated as a (familiar) plane-fitting problem; second, the Sampson error is identical to the geometric error, whereas in the case of a general (non-affine) fundamental matrix (11.9–*p*287) it is only a first-order approximation. As discussed in section 4.2.6-(*p*98) this latter property arises generally with affine (linear) relations because the tangent plane of the Sampson approximation is equivalent to the surface.

14.3.2 The Gold Standard algorithm

Given a set of n corresponding image points $\{\mathbf{x}_i \leftrightarrow \mathbf{x}'_i\}$, we seek the Maximum Likelihood estimate of $\mathtt{F}_\mathrm{A}$ under the assumption that the noise in the image measurements obeys an isotropic, homogeneous Gaussian distribution. This estimate is obtained by minimizing a cost function on geometric image distances:

$$\min_{\{\mathtt{F}_\mathrm{A}, \hat{\mathbf{x}}_i, \hat{\mathbf{x}}'_i\}} \sum_i d(\mathbf{x}_i, \hat{\mathbf{x}}_i)^2 + d(\mathbf{x}'_i, \hat{\mathbf{x}}'_i)^2 \tag{14.4}$$

where as usual $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ are the measured correspondences, and $\hat{\mathbf{x}}_i$ and $\hat{\mathbf{x}}'_i$ are estimated "true" correspondences that satisfy $\hat{\mathbf{x}}'^\mathsf{T}_i \mathtt{F}_\mathrm{A} \hat{\mathbf{x}}_i = 0$ exactly for the estimated affine fundamental matrix. The distances are illustrated in figure 14.3. The true correspondences are subsidiary variables that must also be estimated.

As discussed above, and in section 4.2.5(*p*96), minimizing the cost function (14.4)

Downloaded from https://www.cambridge.org/core. University of Illinois at Urbana - Champaign Library, on 22 Sep 2018 at 15:36:38, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811685.019

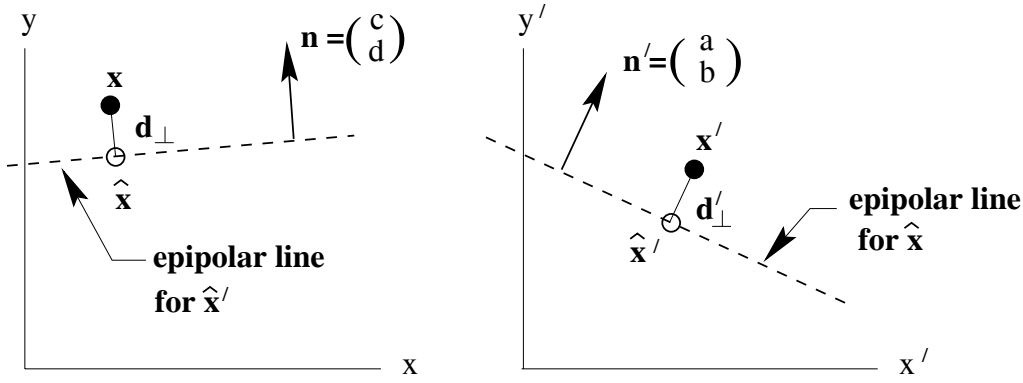


Fig. 14.3. *The MLE of* $\mathtt{F}_{\mathrm{A}}$ *from a set of measured corresponding points* $\{\mathbf{x}_i \leftrightarrow \mathbf{x}'_i\}$ *involves estimating the five parameters* a, b, c, d, e *together with a set of correspondences* $\{\hat{\mathbf{x}}_i \leftrightarrow \hat{\mathbf{x}}'_i\}$ *which exactly satisfy* $\hat{\mathbf{x}}_i'^{\mathsf{T}}\mathtt{F}_{\mathrm{A}}\hat{\mathbf{x}}_i = 0$. *There is a linear solution to this problem.*

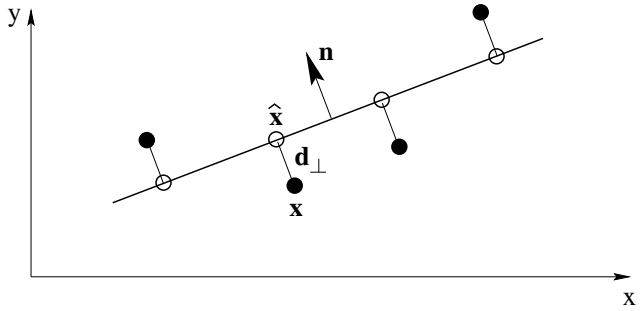


Fig. 14.4. *In 2D a line is the analogue of the hyperplane defined by* $\mathtt{F}_{\mathrm{A}}$, *and the problem of estimating the true correspondence given a measured correspondence, is the problem of determining the closest point* $(\hat{x}, \hat{y})$ *on the line* $ax + by + c$ *to a measurement point* (x, y). *The normal to the line has direction* (a, b), *and the perpendicular distance of the point* (x, y) *from the line is* $d_\perp = (ax + by + c)/\sqrt{a^2 + b^2}$, *so that* $(\hat{x}, \hat{y})^{\mathsf{T}} = (x, y)^{\mathsf{T}} - d_\perp \hat{\mathbf{n}}$, *where* $\hat{\mathbf{n}} = (a, b)/\sqrt{a^2 + b^2}$.

is equivalent to fitting a hyperplane to set of points $\mathbf{X}_i = (x'_i, y'_i, x_i, y_i)^{\mathsf{T}}$ in $\mathbb{R}^4$. The estimated points $\widehat{\mathbf{X}}_i = (\hat{\mathbf{x}}'_i, \hat{\mathbf{y}}'_i, \hat{\mathbf{x}}_i, \hat{\mathbf{y}}_i)^{\mathsf{T}}$ satisfy the equation $\hat{\mathbf{x}}_i'^{\mathsf{T}}\mathtt{F}_{\mathrm{A}}\hat{\mathbf{x}}_i = 0$ which may be written as $(\widehat{\mathbf{X}}_i^{\mathsf{T}}, 1)\mathbf{f} = 0$, where $\mathbf{f} = (a, b, c, d, e)^{\mathsf{T}}$. This is the equation of a point in $\mathbb{R}^4$ on the plane $\mathbf{f}$. We seek the plane $\mathbf{f}$ which minimizes the squared distance between the measured and estimated points, and consequently which minimizes the sum of squared perpendicular distances to the points $\mathbf{X}_i = (x'_i, y'_i, x_i, y_i)^{\mathsf{T}}$.

Geometrically the solution is very simple, and an analogue for line fitting in 2D is illustrated in figure 14.4. The perpendicular distance of a point $\mathbf{X}_i = (x_i, y_i, x'_i, y'_i)^{\mathsf{T}}$ from the plane $\mathbf{f}$ is

$$d_\perp(\mathbf{X}_i, \mathbf{f}) = \frac{ax'_i + by'_i + cx_i + dy_i + e}{\sqrt{a^2 + b^2 + c^2 + d^2}}.$$

Then the matrix $\mathtt{F}_{\mathrm{A}}$ which minimizes (14.4) is determined by minimizing the cost function

$$\mathcal{C} = \sum_i d_\perp(\mathbf{X}_i, \mathbf{f})^2 = \frac{1}{a^2 + b^2 + c^2 + d^2} \sum_i (ax'_i + by'_i + cx_i + dy_i + e)^2 \qquad (14.5)$$

Downloaded from https://www.cambridge.org/core. University of Illinois at Urbana - Champaign Library, on 22 Sep 2018 at 15:36:38, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811685.019

Objective

Given $n \geq 4$ image point correspondences $\{\mathbf{x}_i \leftrightarrow \mathbf{x}'_i\},\ i = 1, \ldots, n$, determine the Maximum Likelihood estimate $\mathtt{F}_{\mathrm{A}}$ of the affine fundamental matrix.

Algorithm

A correspondence is represented as $\mathbf{X}_i = (x'_i, y'_i, x_i, y_i)^{\mathsf{T}}$.

(i) Compute the centroid $\overline{\mathbf{X}} = \frac{1}{n}\sum_i \mathbf{X}_i$ and centre the vectors $\Delta\mathbf{X}_i = \mathbf{X}_i - \overline{\mathbf{X}}$.
(ii) Compute the $n \times 4$ matrix $\mathtt{A}$ with rows $\Delta\mathbf{X}_i^{\mathsf{T}}$.
(iii) Then $\mathbf{N} = (a, b, c, d)^{\mathsf{T}}$ is the singular vector corresponding to the smallest singular value of $\mathtt{A}$, and $e = -\mathbf{N}^{\mathsf{T}}\overline{\mathbf{X}}$. The matrix $\mathtt{F}_{\mathrm{A}}$ has the form (14.1).

Algorithm 14.1. *The Gold Standard algorithm for estimating $\mathtt{F}_{\mathrm{A}}$ from image correspondences.*

over the 5 parameters $\{a, b, c, d, e\}$ of $\mathbf{f}$. Writing $\mathbf{N} = (a, b, c, d)^{\mathsf{T}}$ for the normal to the hyperplane then

$$\mathcal{C} = \frac{1}{\|\mathbf{N}\|^2}\sum_i \left(\mathbf{N}^{\mathsf{T}}\mathbf{X}_i + e\right)^2.$$

This cost function can be minimized by a very simple linear algorithm, equivalent to the classical problem of orthogonal regression to a plane. There are two steps:

The first step is to minimize $\mathcal{C}$ over the parameter e. We obtain

$$\frac{\partial \mathcal{C}}{\partial e} = \frac{1}{\|\mathbf{N}\|^2}\sum_i 2(\mathbf{N}^{\mathsf{T}}\mathbf{X}_i + e) = 0$$

and hence

$$e = -\frac{1}{n}\sum_i (\mathbf{N}^{\mathsf{T}}\mathbf{X}_i) = -\mathbf{N}^{\mathsf{T}}\overline{\mathbf{X}}$$

so the solution hyperplane passes through the data centroid $\overline{\mathbf{X}}$. Substituting for e in the cost function reduces $\mathcal{C}$ to

$$\mathcal{C} = \frac{1}{\|\mathbf{N}\|^2}\sum_i \left(\mathbf{N}^{\mathsf{T}}\Delta\mathbf{X}_i\right)^2$$

where $\Delta\mathbf{X}_i = \mathbf{X}_i - \overline{\mathbf{X}}$ is the vector $\mathbf{X}_i$ relative to the data centroid $\overline{\mathbf{X}}$.

The second step is to minimize this reduced cost function over $\mathbf{N}$. Writing $\mathtt{A}$ for the matrix with rows $\Delta\mathbf{X}_i^{\mathsf{T}}$, it is evident that

$$\mathcal{C} = \|\mathtt{A}\mathbf{N}\|^2/\|\mathbf{N}\|^2.$$

Minimizing this expression is equivalent to minimizing $\|\mathtt{A}\mathbf{N}\|$ subject to $\|\mathbf{N}\| = 1$, which is our usual homogeneous minimization problem solved by the SVD. These steps are summarized in algorithm 14.1.

It is worth noting that the Gold Standard algorithm produces an identical estimate of $\mathtt{F}_{\mathrm{A}}$ to that obtained by the factorization algorithm 18.1(*p*437) for an affine reconstruction from n point correspondences.

Downloaded from https://www.cambridge.org/core. University of Illinois at Urbana - Champaign Library, on 22 Sep 2018 at 15:36:38, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811685.019

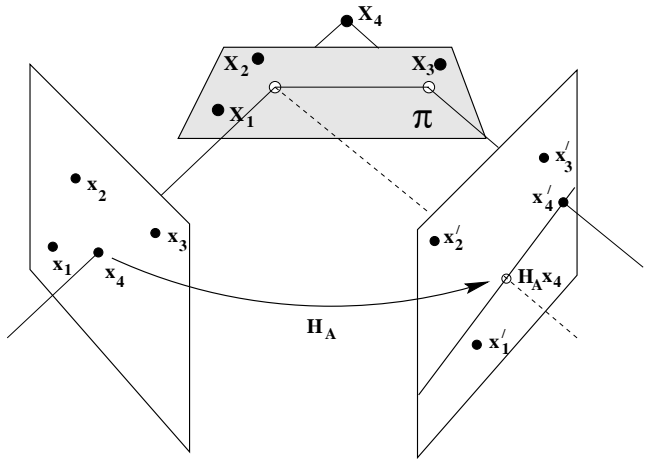


Fig. 14.5. *Computing the affine epipolar line for a minimal configuration of four points. The line is computed by the virtual parallax induced by the plane* $\boldsymbol{\pi}$. *Compare figure 13.7(*p*335).*

Objective

Given four image point correspondences $\{\mathbf{x}_i \leftrightarrow \mathbf{x}'_i\},\ i = 1, \ldots, 4$, compute the affine fundamental matrix.

Algorithm

The first three 3-space points $\mathbf{X}_i,\ i = 1, \ldots, 3$ define a plane $\boldsymbol{\pi}$. See figure 14.5.

(i) Compute the affine transformation matrix $\mathtt{H}_\mathrm{A}$, such that $\mathbf{x}'_i = \mathtt{H}_\mathrm{A}\mathbf{x}_i,\ i = 1, \ldots, 3$.
(ii) Determine the epipolar line in the second view from $\mathbf{l}' = (\mathtt{H}_\mathrm{A}\mathbf{x}_4) \times \mathbf{x}'_4$. The epipole $\mathbf{e}' = (-l'_2, l'_1, 0)^\mathsf{T}$.
(iii) Then for any point $\mathbf{x}$ the epipolar line in the second view is $\mathbf{e}' \times (\mathtt{H}_\mathrm{A}\mathbf{x}) = \mathtt{F}_\mathrm{A}\mathbf{x}$. Hence $\mathtt{F}_\mathrm{A} = [(-l'_2, l'_1, 0)^\mathsf{T}]_\times \mathtt{H}_\mathrm{A}$.

Algorithm 14.2. *The computation of* $\mathtt{F}_\mathrm{A}$ *for a minimal configuration of four point correspondences.*

14.3.3 The minimal configuration

We return to the minimal configuration for estimating $\mathtt{F}_\mathrm{A}$, namely the corresponding images of four points in 3-space in general position. A geometric computation method of $\mathtt{F}_\mathrm{A}$ for this configuration is described in algorithm 14.2. This minimal solution is useful in the case of robust estimation algorithms, such as RANSAC, and will be used here to illustrate degenerate configurations. Note that for this minimal configuration an exact solution is obtained for $\mathtt{F}_\mathrm{A}$, and the linear algorithm of section 14.3.1, the Gold Standard algorithm 14.1, and the minimal algorithm 14.2 give an identical result.

General position. The configuration of four points shown in figure 14.5 demonstrates the conditions necessary for general position of the 3-space points when computing $\mathtt{F}_\mathrm{A}$. Configurations for which $\mathtt{F}_\mathrm{A}$ cannot be computed are degenerate. These fall into two classes: first, degenerate configurations depending only on the structure, for example if the four points are coplanar (so there is no parallax), or if the first three points are collinear (so that $\mathtt{H}_\mathrm{A}$ can't be computed); second, those degeneracies which depend

Downloaded from https://www.cambridge.org/core. University of Illinois at Urbana - Champaign Library, on 22 Sep 2018 at 15:36:38, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811685.019

only on the cameras, for example if the two cameras have the same viewing direction (and so have common centres on the plane at infinity).

Note once again the importance of parallax – as the point $\mathbf{X}_4$ approaches the plane defined by the other three points in figure 14.5 the parallax vector, which determines the epipolar line direction, is monotonically reduced in length. Consequently, the accuracy of the line direction is correspondingly reduced. This result for the minimal configuration is true also of the Gold Standard algorithm 14.1: as relief reduces to zero, i.e. the point set approaches coplanarity, the covariance of the estimated $\mathtt{F}_{\mathrm{A}}$ will increase.

14.4 Triangulation

Suppose we have a measured correspondence $(x, y) \leftrightarrow (x', y')$ and the affine fundamental matrix $\mathtt{F}_{\mathrm{A}}$. We wish to determine the Maximum Likelihood estimate of the true correspondence, $(\hat{x}, \hat{y}) \leftrightarrow (\hat{x}', \hat{y}')$, under the usual assumption that image measurement error is Gaussian. The 3D point may then be determined from the ML estimate correspondence.

As we have seen earlier in chapter 12, the MLE involves determining a "true" correspondence which exactly obeys the affine epipolar geometry, i.e. $(\hat{x}', \hat{y}', 1)\mathtt{F}_{\mathrm{A}}(\hat{x}, \hat{y}, 1)^{\mathsf{T}} = 0$, and also minimizes the image distance to the measured points,

$$(x - \hat{x})^2 + (y - \hat{y})^2 + (x' - \hat{x}')^2 + (y' - \hat{y}')^2.$$

Geometrically the solution is very simple, and is illustrated in 2D in figure 14.4. We seek the closest point on the hyperplane defined by $\mathtt{F}_{\mathrm{A}}$ to the measured correspondence $\mathbf{X} = (x', y', x, y)^{\mathsf{T}}$ in $\mathbb{R}^4$. Again, the Sampson correction (4.11–*p*99) is exact in this case. Algebraically, the normal to the hyperplane has direction $\mathbf{N} = (a, b, c, d)^{\mathsf{T}}$ and the perpendicular distance of a point $\mathbf{X}$ to the hyperplane is given by $d_\perp = (\mathbf{N}^{\mathsf{T}}\mathbf{X} + e)/\|\mathbf{N}\|$, so that

$$\widehat{\mathbf{X}} = \mathbf{X} - d_\perp \frac{\mathbf{N}}{\|\mathbf{N}\|}$$

or in its full detail

$$\begin{pmatrix} \hat{x}' \\ \hat{y}' \\ \hat{x} \\ \hat{y} \end{pmatrix} = \begin{pmatrix} x' \\ y' \\ x \\ y \end{pmatrix} - \frac{(ax' + by' + cx + dy + e)}{(a^2 + b^2 + c^2 + d^2)} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}.$$

14.5 Affine reconstruction

Suppose we have $n \geq 4$ image point correspondences $\mathbf{x}_i \leftrightarrow \mathbf{x}'$, $i = 0, \ldots, n-1$, which for the moment will be assumed noise-free, then we may compute a reconstruction of the 3D points and cameras. In the case of projective cameras (with $n \geq 7$ points) the reconstruction was projective. In the affine case, not surprisingly, the reconstruction is *affine*. We will now give a simple constructive derivation of this result.

An affine coordinate frame in 3-space may be specified by four finite non-coplanar basis points $\mathbf{X}_i$, $i = 0, \ldots, 3$. As illustrated in figure 14.6 one point $\mathbf{X}_0$ is chosen as the

Downloaded from https://www.cambridge.org/core. University of Illinois at Urbana - Champaign Library, on 22 Sep 2018 at 15:36:38, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811685.019

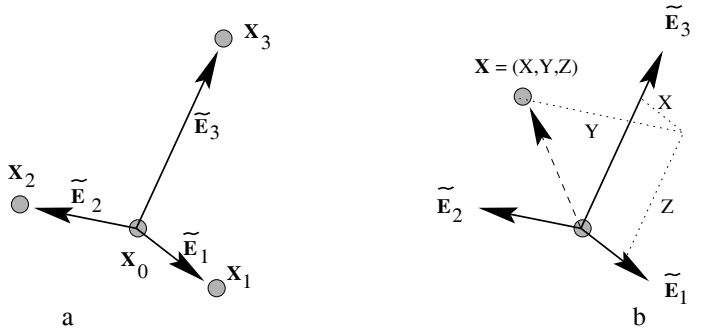


Fig. 14.6. **Affine coordinates.** *(a) four non-coplanar points in 3-space ($\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_3$ and origin $\mathbf{X}_0$) define a set of axes in terms of which other points $\mathbf{X}$ can be assigned affine coordinates $(\mathrm{X}, \mathrm{Y}, \mathrm{Z})$. (b) Each affine coordinate is defined by a ratio of lengths in parallel directions (which is an affine invariant). For example,* X *may be computed by the following two operations: first, $\mathbf{X}$ is projected parallel to $\widetilde{\mathbf{E}}_2$ onto the plane spanned by $\widetilde{\mathbf{E}}_1$ and $\widetilde{\mathbf{E}}_3$. Second, this projected point is projected parallel to $\widetilde{\mathbf{E}}_3$ onto the $\widetilde{\mathbf{E}}_1$ axis. The value of the coordinate* X *is the ratio of the length from the origin of this final projected point to the length of $\widetilde{\mathbf{E}}_1$.*

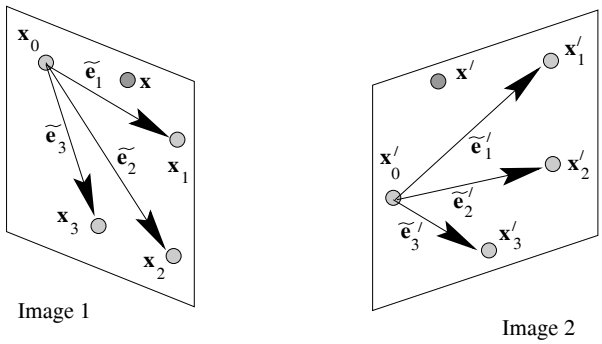


Fig. 14.7. **Reconstruction from two images.** *The affine coordinates of the 3D point $\mathbf{X}$ with image $\mathbf{x}, \mathbf{x}'$ in two views may be computed linearly from the projection of the basis points $\mathbf{x}_i$ and basis vectors $\tilde{\mathbf{e}}_i$ of figure 14.6.*

origin, and the three other points then define basis vectors $\widetilde{\mathbf{E}}_i = \widetilde{\mathbf{X}}_i - \widetilde{\mathbf{X}}_0, i = 1, \ldots, 3$, where $\widetilde{\mathbf{X}}_i$ is the inhomogeneous 3-vector corresponding to $\mathbf{X}_i$. The position of a point $\mathbf{X}$ may then be specified by simple vector addition as

$$\widetilde{\mathbf{X}} = \widetilde{\mathbf{X}}_0 + \mathrm{X}\widetilde{\mathbf{E}}_1 + \mathrm{Y}\widetilde{\mathbf{E}}_2 + \mathrm{Z}\widetilde{\mathbf{E}}_3$$

and $(\mathrm{X}, \mathrm{Y}, \mathrm{Z})$ are the affine coordinates of $\widetilde{\mathbf{X}}$ with respect to this basis. This means that the basis points $\mathbf{X}_i$ have the canonical coordinates $(\mathrm{X}, \mathrm{Y}, \mathrm{Z})^\mathsf{T}$:

$$\widetilde{\mathbf{X}}_0 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad \widetilde{\mathbf{X}}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad \widetilde{\mathbf{X}}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad \widetilde{\mathbf{X}}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \tag{14.6}$$

Given the affine projection of the four basis points in two views, the 3D affine coordinates of any other point can be directly recovered from its image, as will now be demonstrated (see figure 14.7).

Downloaded from https://www.cambridge.org/core. University of Illinois at Urbana - Champaign Library, on 22 Sep 2018 at 15:36:38, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811685.019

Projection with an affine camera may be represented as (6.26–*p*172)

$$\tilde{\mathbf{x}} = \mathtt{M}_{2\times 3}\tilde{\mathbf{X}} + \tilde{\mathbf{t}}$$

where $\tilde{\mathbf{x}} = (x, y)^\mathsf{T}$ is the inhomogeneous 2-vector corresponding to $\mathbf{x}$. Differences of vectors eliminate $\tilde{\mathbf{t}}$. For example the basis vectors project as $\tilde{\mathbf{e}}_i = \mathtt{M}_{2\times 3}\tilde{\mathbf{E}}_i,\ i = 1, \ldots, 3$. Consequently for any point $\mathbf{X}$, its image in the first view is

$$\tilde{\mathbf{x}} - \tilde{\mathbf{x}}_0 \quad = \quad \mathtt{X}\tilde{\mathbf{e}}_1 + \mathtt{Y}\tilde{\mathbf{e}}_2 + \mathtt{Z}\tilde{\mathbf{e}}_3 \tag{14.7}$$

and similarly the image ($\tilde{\mathbf{x}}' = \mathtt{M}'_{2\times 3}\tilde{\mathbf{X}} + \tilde{\mathbf{t}}'$) in the second view is

$$\tilde{\mathbf{x}}' - \tilde{\mathbf{x}}'_0 \quad = \quad \mathtt{X}\tilde{\mathbf{e}}'_1 + \mathtt{Y}\tilde{\mathbf{e}}'_2 + \mathtt{Z}\tilde{\mathbf{e}}'_3. \tag{14.8}$$

Each equation (14.7) and (14.8) imposes two linear constraints on the unknown affine coordinates $\mathtt{X}, \mathtt{Y}, \mathtt{Z}$ of the space point $\mathbf{X}$. All the other terms in the equations are known from image measurements (for example the image basis vectors $\tilde{\mathbf{e}}_i, \tilde{\mathbf{e}}'_i$ are computed from the projection of the four basis points $\tilde{\mathbf{X}}_i,\ i = 0, \ldots, 3$). Thus, there are four linear simultaneous equations in the three unknowns $\mathtt{X}, \mathtt{Y}, \mathtt{Z}$, and the solution is straightforward. This demonstrates that the affine coordinates of a point $\mathbf{X}$ may be computed from its image in two views.

The cameras for the two views, $\mathtt{P}_\mathrm{A}, \mathtt{P}'_\mathrm{A}$, may be computed from the correspondences between the 3-space points $\tilde{\mathbf{X}}_i$, with coordinates given in (14.6), and their measured images. For example, $\mathtt{P}_\mathrm{A}$ is computed from the correspondence $\tilde{\mathbf{x}}_i \leftrightarrow \tilde{\mathbf{X}}_i,\ i = 0, \ldots, 3$.

The above development is not optimal, because the basis points are treated as exact, and all measurement error associated with the fifth point $\mathbf{X}$. An optimal reconstruction algorithm, where reprojection error is minimized over all points, is very straightforward in the affine case. However, its description is postponed until section 18.2(*p*436) because the factorization algorithm described there is applicable to any number of views.

Example 14.2. Affine reconstruction
A 3D reconstruction is computed for the hole punch images of figure 14.2 by choosing four points as the affine basis, and then computing the affine coordinates of each of the remaining points in turn by the linear method above. Two views of the resulting reconstruction are shown in figure 14.8. Note, however, that this five-point method is *not* recommended. Instead the optimal affine reconstruction algorithm 18.1(*p*437) should be used. △

14.6 Necker reversal and the bas-relief ambiguity

We have seen in the previous section that in the absence of any calibration information, an affine reconstruction is obtained from point correspondences alone. In this section we show that even if the camera calibration is known there remains a family of reconstruction ambiguities which cannot be resolved in the two-view case.

This situation differs from that of perspective projection where once the internal calibration is determined the camera motion is determined up to a finite number of

Downloaded from https://www.cambridge.org/core. University of Illinois at Urbana - Champaign Library, on 22 Sep 2018 at 15:36:38, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811685.019

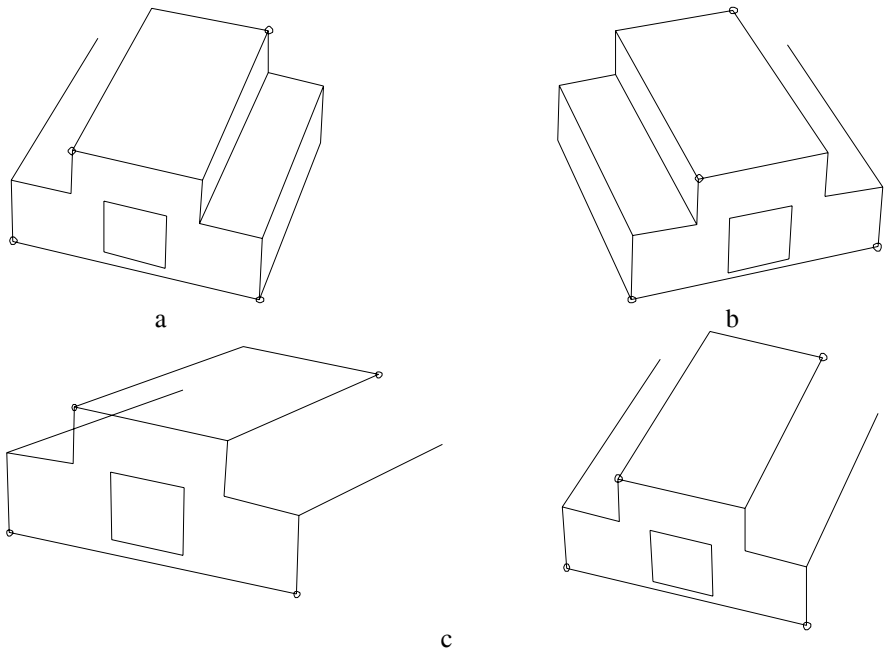


Fig. 14.8. **Affine reconstruction.** *(a)(b) Wireframe outline of the hole punch from the two images of figure 14.2. The circles show the points selected as the affine basis. The lines are for visualization only. (c) Two views of the 3D affine structure computed from the vertices of the wireframe.*

ambiguities (from the essential matrix, see section 9.6(*p*257)). For parallel projection there are two important additional ambiguities: a finite reflection ambiguity (Necker reversal); and a one-parameter family rotation ambiguity (the bas-relief ambiguity).

Necker reversal ambiguity. This arises because an object rotating by ρ and its mirror image rotating by $-\rho$ generate the same image under parallel projection, see figure 14.9(a). Thus, structure is only recovered up to a reflection about the frontal plane. This ambiguity is absent in the perspective case because the points have different depths in the two interpretations and so do not project to coincident image points.

The bas-relief ambiguity. This is illustrated in figure 14.9(b). Imagine a set of parallel rays from one camera, and consider adjusting a set of parallel rays from a second camera until each ray intersects its corresponding ray. The rays lie in a family of parallel epipolar planes, and there remains the freedom to rotate one camera about the normal to these planes whilst maintaining incidence of the rays. This *bas-relief* (or *depth–turn*) ambiguity is a one-parameter family of solutions for the rotation angle and depth. The parameters of depth, ΔZ, and rotation, $\sin\rho$, are confounded and cannot be determined individually – only their product can be computed. Consequently, a shallow object experiencing a large turn (i.e. small ΔZ and large ρ) generates the same image as a deep object experiencing a small turn (i.e. large ΔZ and small ρ). The name arises from bas-relief sculptures. Fixing the depth or the angle ρ determines the structure and the motion uniquely. Extra points cannot resolve this ambiguity, but an additional view (i.e. three views) will in general resolve it.

Downloaded from https://www.cambridge.org/core. University of Illinois at Urbana - Champaign Library, on 22 Sep 2018 at 15:36:38, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811685.019

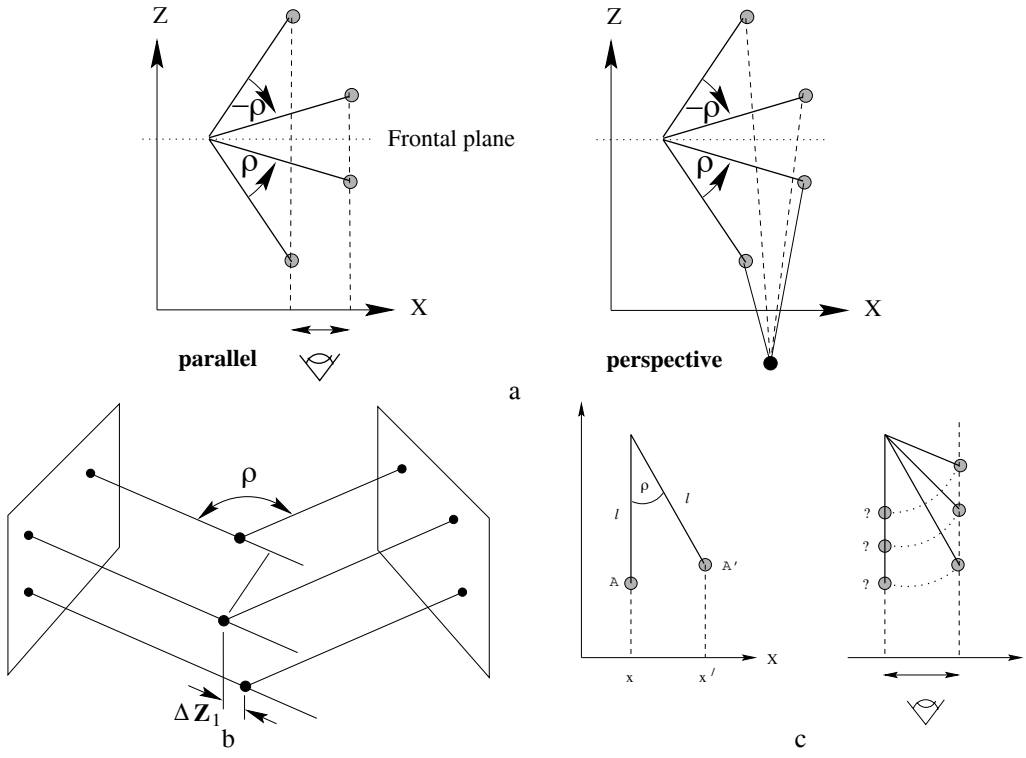


Fig. 14.9. **Motion ambiguities under parallel projection.** *(a) Necker reversal: a rotating object generates the same image as its mirror object rotating in the opposite sense. Under perspective projection the images are different. (b) The cameras can rotate (by ρ) and still preserve ray intersections. This cannot happen for perspective cameras. (c) The bas-relief ambiguity: consider a rod of length l, which rotates through an angle ρ. That is $x' - x = l \sin \rho$. This* bas–relief *(or* depth–turn*) ambiguity is so-named because a shallow object experiencing a large turn (i.e. small l and big ρ) generates the same image as a deep object experiencing a small turn (i.e. large l and small ρ).*

This ambiguity casts light on the stability of reconstruction from two perspective cameras: as imaging conditions approach affine the rotation angle will be poorly estimated, but the product of the rotation angle and depth will be stable.

14.7 Computing the motion

In this section expressions for computing the camera motion from $\mathtt{F}_{\mathrm{A}}$ will be given for the case of two weak perspective cameras (section 6.3.4(*p*170)). These cameras may be chosen as

$$\mathtt{P} = \begin{bmatrix} \alpha_x & & \\ & \alpha_y & \\ & & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \mathtt{P}' = \begin{bmatrix} \alpha'_x & & \\ & \alpha'_y & \\ & & 1 \end{bmatrix} \begin{bmatrix} \mathbf{r}^{1\mathsf{T}} & t_1 \\ \mathbf{r}^{2\mathsf{T}} & t_2 \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix}$$

where $\mathbf{r}_1$ and $\mathbf{r}_2$ are the first and second rows of the rotation matrix $\mathtt{R}$ between the views. We will assume that the aspect ratio α_y/α_x is known in both cameras, but that the relative scaling $s = \alpha'_x/\alpha_x$ is unknown. $s > 1$ for a "looming" object and $s < 1$ for one that is "receding". As has been seen, the complete rotation $\mathtt{R}$ cannot be computed from two weak perspective views, resulting in the *bas-relief* ambiguity. Nevertheless

Downloaded from https://www.cambridge.org/core. University of Illinois at Urbana - Champaign Library, on 22 Sep 2018 at 15:36:38, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811685.019

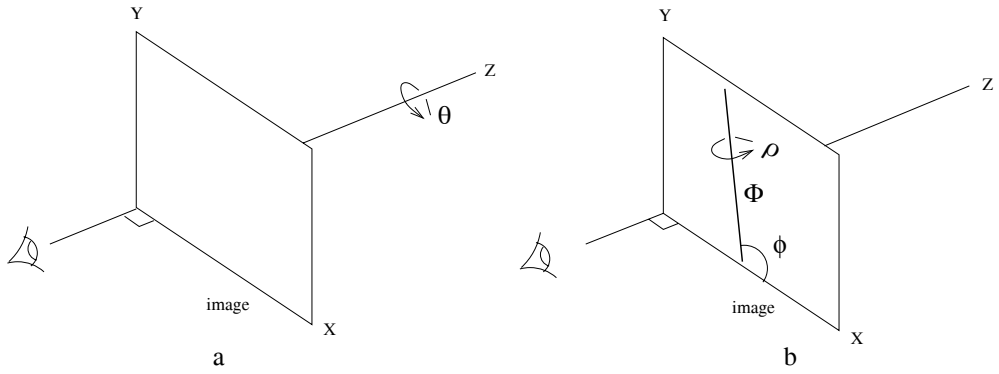


Fig. 14.10. **The rotation representation.** *(a) rotation by θ about the* Z*-axis; (b) subsequent rotation by ρ about a fronto–parallel axis $\mathbf{\Phi}$ angled at ϕ to the* X*-axis. The $\mathbf{\Phi}$-axis has components* $(\cos\phi, \sin\phi, 0)^\mathsf{T}$.

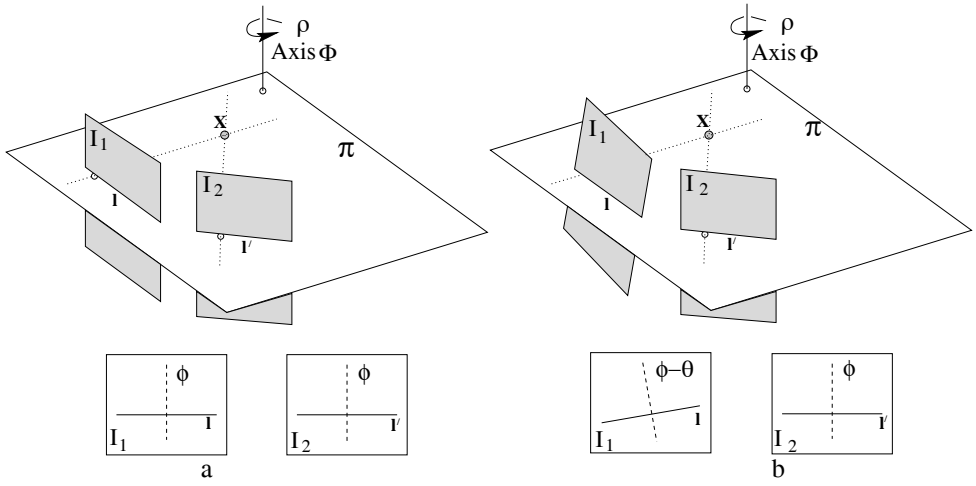


Fig. 14.11. *The camera rotates about the axis $\mathbf{\Phi}$ which is parallel to the image plane. The intersection of the epipolar plane $\boldsymbol{\pi}$ with the image planes gives epipolar lines* $\mathbf{l}$ *and* $\mathbf{l}'$*, and the projections of $\mathbf{\Phi}$ in the images are orthogonal to these epipolar lines: (a) no cyclotorsion occurs ($\theta = 0°$); (b) the camera counter-rotates by θ in I_1, so the orientation of the epipolar lines changes by θ.*

the remaining motion parameters can be computed from $\mathtt{F}_\mathrm{A}$, and their computation is straightforward.

To represent the motion we will use a rotation representation introduced by Koenderink and van Doorn [Koenderink-91]. As will be seen, this has the advantage that it isolates the parameter ρ of the bas-relief ambiguity, which cannot be computed from the affine epipolar geometry. In this representation the rotation $\mathtt{R}$ between the views is decomposed into two rotations (see figure 14.10),

$$\mathtt{R} = \mathtt{R}_\rho \, \mathtt{R}_\theta. \tag{14.9}$$

First, there is a cyclo-rotation $\mathtt{R}_\theta$ in the image plane through an angle θ (i.e. about the line of sight). This is followed by a rotation $\mathtt{R}_\rho$ through an angle ρ about an axis $\mathbf{\Phi}$ with direction parallel to the image plane, and angled at ϕ to the positive X-axis, i.e. a pure rotation *out of* the image plane.

Downloaded from https://www.cambridge.org/core. University of Illinois at Urbana - Champaign Library, on 22 Sep 2018 at 15:36:38, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811685.019

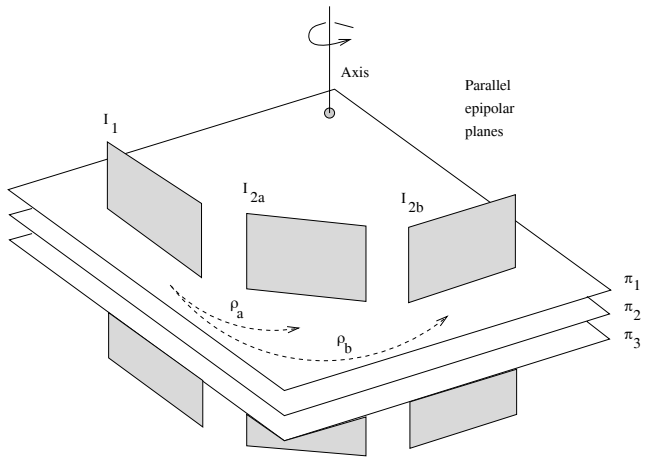


Fig. 14.12. *The scene can be sliced into parallel epipolar planes. The magnitude of ρ has no effect on the epipolar geometry (provided $\rho \neq 0$), so it is indeterminate from two views.*

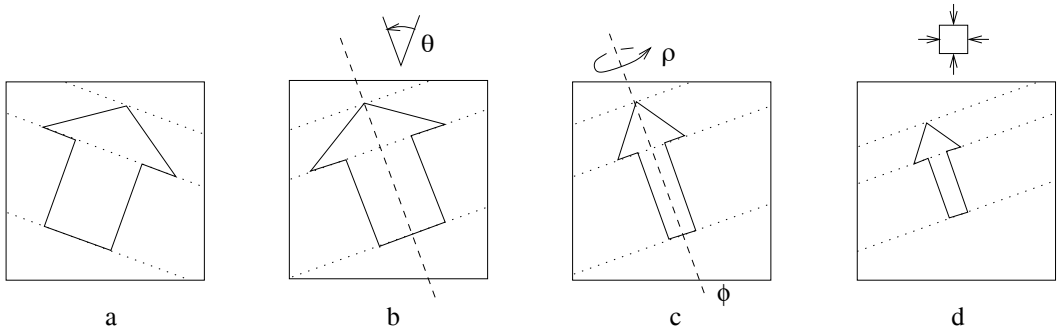


Fig. 14.13. *The effect of scale and rotation angles on the epipolar lines for an object moving relative to a stationary camera. This also illustrates the assumed sequence of events accounting for the transition from I_1 to I_2: (a) I_1; (b) cyclotorsion (θ); (c) rotation out of the plane (ϕ and ρ); (d) scaling, giving I_2.*

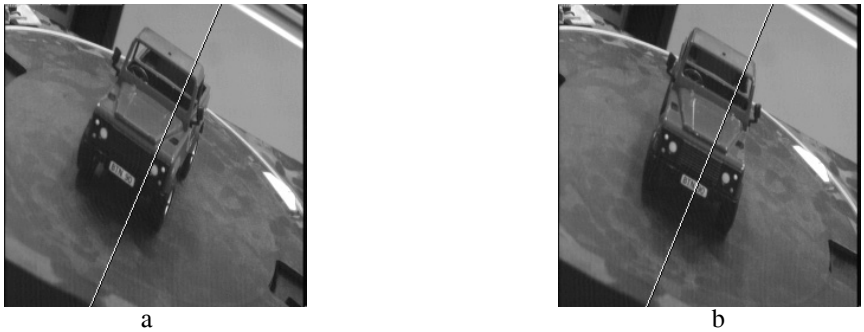
Solving for s, ϕ and θ. It is now shown that the scale factor (s), the projection of the axis of rotation (ϕ) and the cyclo-rotation angle (θ) may be computed directly from the affine epipolar geometry. The solution is preceded by a geometric explanation of how the epipolar lines relate to the unknown motion parameters.

Consider a camera rotating about an axis $\mathbf{\Phi}$ lying parallel to the image plane (figure 14.11(a)). The epipolar plane $\boldsymbol{\pi}$ is perpendicular to both this axis and the two images, and intersects the images in the epipolar lines $\mathbf{l}$ and $\mathbf{l}'$. Consequently:

- *The projection of the axis of rotation $\mathbf{\Phi}$ is perpendicular to the epipolar lines.*

This relation still holds if there is additionally a cyclotorsion θ in the image plane (figure 14.11(b)); the axis $\mathbf{\Phi}$ and intersection $\mathbf{l}'$ remain fixed in space, and are simply observed at a new angle in the image, maintaining the orthogonality between the epipolar lines and the projected axis. The orientations of the epipolar lines in the two images therefore differ by θ. Importantly, changing the magnitude of the turn angle ρ doesn't alter the epipolar geometry in any way (figure 14.12). This angle is therefore indeterminate from two views, a consequence of the bas-relief ambiguity.

Downloaded from https://www.cambridge.org/core. University of Illinois at Urbana - Champaign Library, on 22 Sep 2018 at 15:36:38, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811685.019



a b

Fig. 14.14. **Computing motion from affine epipolar geometry.** *(a)(b) Two views of a buggy rotating on a turntable. The computed rotation axis is superimposed on the image, drawn to pass through the image centre. The ground truth axis is, of course, perpendicular to the turntable in the world.*

Figure 14.13 illustrates the effect of scale. Consider a 3D object to be sliced into parallel epipolar planes, with each plane constraining how a particular slice of the object moves. Altering the effective size of the object (e.g. by moving closer to it) simply changes the relative spacing between successive epipolar planes.

In summary, cyclotorsion simply rotates the epipolar lines, rotation out of the plane causes foreshortening along the epipolar lines (orthogonal to $\mathbf{\Phi}$), and a scale change uniformly alters the epipolar line spacing (figure 14.13).

It can be shown (and is left as exercise) that s, θ and ϕ can be computed directly from the affine epipolar geometry as

$$\tan\phi = \frac{b}{a}, \quad \tan(\phi - \theta) = \frac{d}{c} \text{ and } s^2 = \frac{c^2 + d^2}{a^2 + b^2}, \tag{14.10}$$

with $s > 0$ (by definition). Note that ϕ is the angle of projection in I_2 of the axis of rotation out of the plane, while $(\phi - \theta)$ is its angle of projection in I_1.

Example 14.3. Motion computed from the affine fundamental matrix
figure 14.14 shows two images of a buggy rotating on a turntable. The image is 256×256 pixels with an aspect ratio of 0.65. The affine fundamental matrix is computed using algorithm 14.1, and the motion parameters computed from $\mathtt{F}_\mathrm{A}$ using (14.10) above. The computed rotation axis is superimposed on the image. △

14.8 Closure

14.8.1 The literature

Koenderink and van Doorn [Koenderink-91] set the scene for affine reconstruction from two affine cameras. This paper should be read by all. The affine fundamental matrix was first defined in [Zisserman-92]. The computation of the motion parameters from $\mathtt{F}_\mathrm{A}$ is described in Shapiro *et al.* [Shapiro-95], and in particular the cases where a third view does not resolve the bas–relief ambiguity. A helpful eigenvector analysis of the ambiguity is given in [Szeliski-96]. The three view affine motion case is treated quite elegantly in [Shimshoni-99].

Downloaded from https://www.cambridge.org/core. University of Illinois at Urbana - Champaign Library, on 22 Sep 2018 at 15:36:38, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811685.019

14.8.2 Notes and exercises

(i) A scene plane induces an *affine* transformation between two affine cameras. There is a three-parameter family of such affinities defined by the three-parameter family of planes in $\mathbb{R}^3$. Given $\mathtt{F}_\mathrm{A}$, this family of affinities may be written as (result 13.3(*p*328)) $\mathtt{H}_\mathrm{A} = [\mathbf{e}']_\times \mathtt{F}_\mathrm{A} + \mathbf{e}'\mathbf{v}^\mathsf{T}$, where $\mathtt{F}_\mathrm{A}{}^\mathsf{T}\mathbf{e}' = \mathbf{0}$, and the 3-vector $\mathbf{v}$ parametrizes the family of planes. Conversely, show that given $\mathtt{H}_\mathrm{A}$, the homography induced by a scene plane, then $\mathtt{F}_\mathrm{A}$ is determined up to a one-parameter ambiguity.

(ii) Consider a perspective camera, i.e. the matrix does not have the affine form. Show that if the camera motion consists of a translation parallel to the image plane, and a rotation about the principal axis, then $\mathtt{F}$ has the affine form. This shows that a fundamental matrix with affine form does *not* imply that the imaging conditions are affine. Are there other camera motions which generate a fundamental matrix with the affine form?

(iii) Two affine cameras, $\mathtt{P}_\mathrm{A}, \mathtt{P}'_\mathrm{A}$, uniquely define an affine fundamental matrix $\mathtt{F}_\mathrm{A}$ by (9.1–*p*244). Show that if the cameras are transformed on the right by a common affine transformation, i.e. $\mathtt{P}_\mathrm{A} \mapsto \mathtt{P}_\mathrm{A}\mathtt{H}_\mathrm{A}, \mathtt{P}'_\mathrm{A} \mapsto \mathtt{P}'_\mathrm{A}\mathtt{H}_\mathrm{A}$, the transformed cameras define the original $\mathtt{F}_\mathrm{A}$. This shows that the affine fundamental matrix is invariant to an affine transformation of the world coordinates.

(iv) Suppose one of the cameras is affine and the other is a perspective camera. Show that in general in this case the epipoles in both views are finite.

(v) The 4×4 permutation homography

$$\mathtt{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

maps the canonical matrix of a finite projective camera $\mathtt{P} = [\mathtt{I} \mid \mathbf{0}]$ into the canonical matrix of parallel projection $\mathtt{P}_\mathrm{A}$:

$$\mathtt{P}_\mathrm{A} = [\mathtt{I} \mid \mathbf{0}]\,\mathtt{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Show, by applying this transformation to a pair of finite projective camera matrices, that the results of this chapter (such as the properties listed in table 14.1-(*p*348)) can be generated directly from their non-affine counterparts of the previous chapters. In particular derive an expression for a pair of affine cameras $\mathtt{P}_\mathrm{A}, \mathtt{P}'_\mathrm{A}$ consistent with $\mathtt{F}_\mathrm{A}$.

Downloaded from https://www.cambridge.org/core. University of Illinois at Urbana - Champaign Library, on 22 Sep 2018 at 15:36:38, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811685.019

Downloaded from https://www.cambridge.org/core. University of Illinois at Urbana - Champaign Library, on 22 Sep 2018 at 15:36:38, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811685.019