Chapter 12

THE GEOMETRY OF MULTIPLE VIEWS

Despite the wealth of information contained in a photograph, the depth of a scene point along the corresponding projection ray is not directly accessible in a single image. With at least two pictures, on the other hand, depth can be measured through triangulation. This is of course one of the reasons why most animals have at least two eyes and/or move their head when looking for friend or foe, as well as the motivation for equipping autonomous robots with stereo or motion analysis systems. Before building such a program, we must understand how several views of the same scene constrain its three-dimensional structure as well as the corresponding camera configurations. This is the goal of this chapter.

In particular, we will elucidate the geometric and algebraic constraints that hold among two, three, or more views of the same scene. In the familiar setting of binocular stereo vision, we will show that the first image of any point must lie in the plane formed by its second image and the optical centers of the two cameras. This *epipolar constraint* can be represented algebraically by a 3×3 matrix called the *essential matrix* when the intrinsic parameters of the cameras are known, and the *fundamental matrix* otherwise. Three pictures of the same line will introduce a different constraint, namely that the intersection of the planes formed by their preimages be degenerate. Algebraically, this geometric relationship can be represented by a $3 \times 3 \times 3$ *trifocal tensor*. More images will introduce additional constraints, for example four projections of the same point will satisfy certain quadrilinear relations whose coefficients are captured by the *quadrifocal tensor*, etc. Remarkably, the equations satisfied by multiple pictures of the same scene feature can be set up without any knowledge of the cameras and the scene they observe, and a number of methods for estimating their parameters directly from image data will be presented in this chapter.

Computer vision is not the only scientific field concerned with the geometry of multiple views: the goal of photogrammetry, already mentioned in Chapter 6, is precisely to recover quantitative geometric information from multiple pictures. Applications of the epipolar and trifocal constraints to the classical photogrammetry problem of *transfer* (i.e., the prediction of the position of a point in an image given its position in a number of reference pictures) will be briefly discussed in this chapter, along with some examples. Many more applications in the domains of stereo and motion analysis will be presented in latter chapters.

12.1 Two Views

12.1.1 Epipolar Geometry

Consider the images p and p' of a point P observed by two cameras with optical centers O and O'. These five points all belong to the *epipolar plane* defined by the two intersecting rays OP and $O'P$ (Figure 12.1). In particular, the point p' lies on the line l' where this plane and the retina Π' of the second camera intersect. The line l' is the *epipolar line* associated with the point p, and it passes through the point e' where the *baseline* joining the optical centers O and O' intersects Π'. Likewise, the point p lies on the epipolar line l associated with the point p', and this line passes through the intersection e of the baseline with the plane Π.

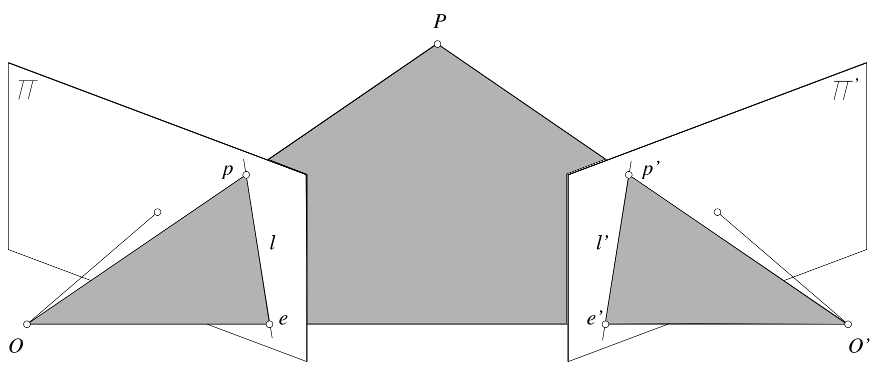


Figure 12.1. Epipolar geometry: the point P, the optical centers O and O' of the two cameras, and the two images p and p' of P all lie in the same plane.

The points e and e' are called the *epipoles* of the two cameras. The epipole e' is the (virtual) image of the optical center O of the first camera in the image observed by the second camera, and vice versa. As noted before, if p and p' are images of the same point, then p' must lie on the epipolar line associated with p. This *epipolar constraint* plays a fundamental role in stereo vision and motion analysis.

Let us assume for example that we know the intrinsic and extrinsic parameters of the two cameras of a stereo rig. We will see in Chapter 13 that the most difficult part of stereo data analysis is establishing correspondences between the two images, i.e., deciding which points in the right picture match the points in the left one. The epipolar constraint greatly limits the search for these correspondences: indeed, since we assume that the rig is calibrated, the coordinates of the point p completely determine the ray joining O and p, and thus the associated epipolar plane $OO'p$ and epipolar line. The search for matches can be restricted to this line instead of the whole image (Figure 12.2). In two-frame motion analysis on the other hand, each camera may be internally calibrated, but the rigid transformation separating the two camera coordinate systems is unknown. In this case, the epipolar geometry obviously constrains the set of possible motions. The next sections explore several variants of this situation.

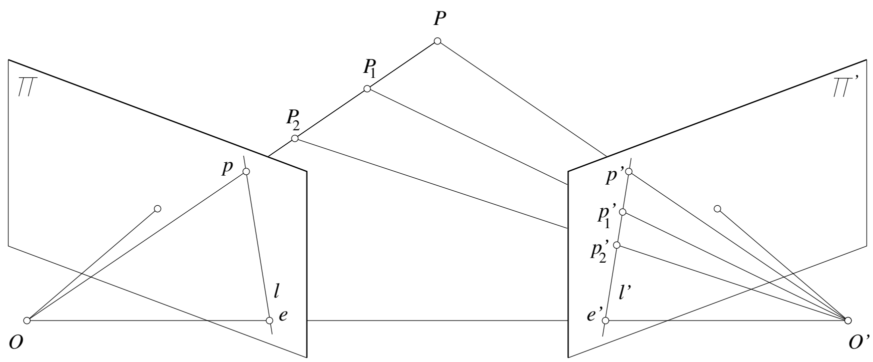


Figure 12.2. Epipolar constraint: given a calibrated stereo rig, the set of possible matches for the point p is constrained to lie on the associated epipolar line l'.

12.1.2 The Calibrated Case

Here we assume that the intrinsic parameters of each camera are known, so $\boldsymbol{p} = \hat{\boldsymbol{p}}$. Clearly, the epipolar constraint implies that the three vectors $\overrightarrow{Op}$, $\overrightarrow{O'p'}$, and $\overrightarrow{OO'}$ are coplanar. Equivalently, one of them must lie in the plane spanned by the other two, or

$$\overrightarrow{Op} \cdot [\overrightarrow{OO'} \times \overrightarrow{O'p'}] = 0.$$

We can rewrite this coordinate-independent equation in the coordinate frame associated to the first camera as

$$\boldsymbol{p} \cdot [\boldsymbol{t} \times (\mathcal{R}\boldsymbol{p}')], \tag{12.1.1}$$

where $\boldsymbol{p} = (u, v, 1)^T$ and $\boldsymbol{p}' = (u', v', 1)^T$ denote the homogenous image coordinate vectors of p and p', $\boldsymbol{t}$ is the coordinate vector of the translation $\overrightarrow{OO'}$ separating the two coordinate systems, and $\mathcal{R}$ is the rotation matrix such that a free vector with coordinates $\boldsymbol{w}'$ in the second coordinate system has coordinates $\mathcal{R}\boldsymbol{w}'$ in the first one (in this case the two projection matrices are given in the coordinate system attached to the first camera by $(\,\mathrm{Id} \quad \boldsymbol{0}\,)$ and $(\,\mathcal{R}^T, -\mathcal{R}^T\boldsymbol{t}\,)$).

Equation (12.1.1) can finally be rewritten as

$$\boldsymbol{p}^T \mathcal{E} \boldsymbol{p}' = 0, \tag{12.1.2}$$

where $\mathcal{E} = [\boldsymbol{t}_\times]\mathcal{R}$, and $[\boldsymbol{a}_\times]$ denotes the skew-symmetric matrix such that $[\boldsymbol{a}_\times]\boldsymbol{x} = \boldsymbol{a}\times\boldsymbol{x}$ is the cross-product of the vectors $\boldsymbol{a}$ and $\boldsymbol{x}$. The matrix $\mathcal{E}$ is called the *essential matrix*, and it was first introduced by Longuet-Higgins [1981]. Its nine coefficients are only defined up to scale, and they can be parameterized by the three degrees of freedom of the rotation matrix $\mathcal{R}$ and the two degrees of freedom defining the direction of the translation vector $\boldsymbol{t}$.

Note that $\mathcal{E}\boldsymbol{p}'$ can be interpreted as the coordinate vector representing the epipolar line associated with the point p' in the first image: indeed, an image line l can be defined by its equation $au + bv + c = 0$, where (u, v) denote the coordinates of a point on the line, (a, b) is the unit normal to the line, and $-c$ is the (signed) distance between the origin and l. Alternatively, we can define the line equation in terms of the homogeneous coordinate vector $\boldsymbol{p} = (u, v, 1)^T$ of a point on the line and the vector $\boldsymbol{l} = (a, b, c)^T$ by $\boldsymbol{l}\cdot\boldsymbol{p} = 0$, in which case the constraint $a^2 + b^2 = 1$ is relaxed since the equation holds independently of any scale change applied to $\boldsymbol{l}$. In this context, (12.1.2) expresses the fact that the point p lies on the epipolar line associated with the vector $\mathcal{E}\boldsymbol{p}'$. By symmetry, it is also clear that $\mathcal{E}^T\boldsymbol{p}$ is the coordinate vector representing the epipolar line associated with p in the second image.

It is obvious that essential matrices are singular since $\boldsymbol{t}$ is parallel to the coordinate vector $\boldsymbol{e}$ of the left epipole, so that $\mathcal{E}^T\boldsymbol{e} = -\mathcal{R}^T[\boldsymbol{t}_\times]\boldsymbol{e} = 0$. Likewise, it is easy to show that $\boldsymbol{e}'$ is a zero eigenvector of $\mathcal{E}$. As shown by Huang and Faugeras [1989], essential matrices are in fact characterized by the fact that they are singular with two equal non-zero singular values (see exercises).

12.1.3 Small Motions

Let us now turn our attention to *infinitesimal* displacements. We consider a moving camera with translational velocity $\boldsymbol{v}$ and rotational velocity $\boldsymbol{\omega}$ and rewrite (12.1.2) for two frames separated by a small time interval δt. Let us denote by $\dot{\boldsymbol{p}} = (\dot{u}, \dot{v}, 0)^T$ the velocity of the point p, or *motion field*. Using the exponential representation of rotations,[1] we have (to first order):

$$\begin{cases} \boldsymbol{t} = \delta t\,\boldsymbol{v}, \\ \mathcal{R} = \mathrm{Id} + \delta t\,[\boldsymbol{\omega}_\times], \\ \boldsymbol{p}' = \boldsymbol{p} + \delta t\,\dot{\boldsymbol{p}}. \end{cases}$$

Substituting in (12.1.2) yields

$$\boldsymbol{p}^T[\boldsymbol{v}_\times](\mathrm{Id} + \delta t\,[\boldsymbol{\omega}_\times])(\boldsymbol{p} + \delta t\,\dot{\boldsymbol{p}}) = 0,$$

and neglecting all terms of order two or greater in δt yields:

$$\boldsymbol{p}^T([\boldsymbol{v}_\times][\boldsymbol{\omega}_\times])\boldsymbol{p} - (\boldsymbol{p}\times\dot{\boldsymbol{p}})\cdot\boldsymbol{v} = 0. \tag{12.1.3}$$

[1] The matrix associated with the rotation whose axis is the unit vector $\boldsymbol{a}$ and whose angle is θ can be shown to be equal to $e^{\theta[\boldsymbol{a}_\times]} \stackrel{\text{def}}{=} \sum_{i=0}^{+\infty} \frac{1}{i!}(\theta[\boldsymbol{a}_\times])^i$.

Equation (12.1.3) is simply the instantaneous form of the Longuet-Higgins relation (12.1.2) which captures the epipolar geometry in the discrete case. Note that in the case of pure translation we have $\boldsymbol{\omega} = 0$, thus $(\boldsymbol{p} \times \dot{\boldsymbol{p}}) \cdot \boldsymbol{v} = 0$. In other words, the three vectors $\boldsymbol{p} = \overrightarrow{op}$, $\dot{\boldsymbol{p}}$ and $\boldsymbol{v}$ must be coplanar. If e denotes the infinitesimal epipole, or *focus of expansion*, i.e., the point where the line passing through the optical center and parallel to the velocity vector $\boldsymbol{v}$ pierces the image plane, we obtain the well known result that the motion field points toward the focus of expansion under pure translational motion (Figure 12.3).

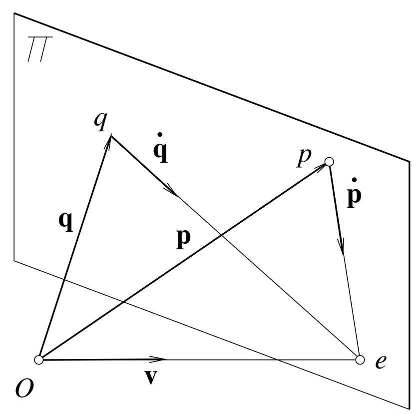


Figure 12.3. Focus of expansion: under pure translation, the motion field at every point in the image points toward the focus of expansion.

12.1.4 The Uncalibrated Case

The Longuet-Higgins relation holds for *internally calibrated* cameras, whose intrinsic parameters are known so that image positions can be expressed in normalized coordinates. When these parameters are unknown (*uncalibrated* cameras), we can write $\boldsymbol{p} = \mathcal{K}\hat{\boldsymbol{p}}$ and $\boldsymbol{p}' = \mathcal{K}'\hat{\boldsymbol{p}}'$, where $\mathcal{K}$ and $\mathcal{K}'$ are 3×3 calibration matrices, and $\hat{\boldsymbol{p}}$ and $\hat{\boldsymbol{p}}'$ are normalized image coordinate vectors. The Longuet-Higgins relation holds for these vectors, and we obtain

$$\boldsymbol{p}^T \mathcal{F} \boldsymbol{p}' = 0, \tag{12.1.4}$$

where the matrix $\mathcal{F} = \mathcal{K}^{-T}\mathcal{E}\mathcal{K}'^{-1}$, called the *fundamental matrix*, is not, in general, an essential matrix.[2] It has again rank two, and the eigenvector of $\mathcal{F}$ (resp. $\mathcal{F}^T$) corresponding to its zero eigenvalue is as before the position $\boldsymbol{e}'$ (resp. $\boldsymbol{e}$) of the epipole. Note that $\mathcal{F}\boldsymbol{p}'$ (resp. $\mathcal{F}^T\boldsymbol{p}$) represents the epipolar line corresponding to the point $\boldsymbol{p}'$ (resp. $\boldsymbol{p}$) in the first (resp. second) image.

[2]Small motions can also be handled in the uncalibrated setting. In particular, Viéville and Faugeras [1995] have derived an equation similar to (12.1.3) that characterizes the motion field of a camera with varying intrinsic parameters.

The rank-two constraint means that the fundamental matrix only admits seven independent parameters. Several choices of parameterization are possible, but the most natural one is in terms of the coordinate vectors $\boldsymbol{e} = (\alpha, \beta)^T$ and $\boldsymbol{e}' = (\alpha', \beta')^T$ of the two epipoles, and of the so-called *epipolar transformation* that maps one set of epipolar lines onto the other one. We will examine the properties of the epipolar transformation in Chapter 15 in the context of structure from motion. For the time being, let us just note (without proof) that this transformation is parameterized by four numbers a, b, c, d, and that the fundamental matrix can be written as

$$\mathcal{F} = \begin{pmatrix} b & a & -a\beta - b\alpha \\ -d & -c & c\beta + d\alpha \\ d\beta' - b\alpha' & c\beta' - a\alpha' & -c\beta\beta' - d\beta'\alpha + a\beta\alpha' + b\alpha\alpha' \end{pmatrix}. \tag{12.1.5}$$

12.1.5 Weak Calibration

As mentioned earlier, the essential matrix is defined up to scale by five independent parameters. It is therefore possible (at least in principle), to calculate it by writing (12.1.2) for five point correspondences. Likewise, the fundamental matrix is defined by seven independent coefficients (the parameters a, b, c, d in (12.1.5) are only defined up to scale) and can in principle be estimated from seven point correspondences. Methods for estimating the essential and fundamental matrices from a minimal number of parameters indeed exist (see [Faugeras, 1993] and Section 12.4), but they are far too involved to be described here. This section addresses the simpler problem of estimating the epipolar geometry from a redundant set of point correspondences between two images taken by cameras with unknown intrinsic parameters, a process known as *weak calibration*.

Note that the epipolar constraint (12.1.4) is a linear equation in the nine coefficients of the fundamental matrix $\mathcal{F}$:

$$(u, v, 1) \begin{pmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{pmatrix} \begin{pmatrix} u' \\ v' \\ 1 \end{pmatrix} = 0 \Leftrightarrow (uu', uv', u, vu', vv', v, u', v', 1) \begin{pmatrix} F_{11} \\ F_{12} \\ F_{13} \\ F_{21} \\ F_{22} \\ F_{23} \\ F_{31} \\ F_{32} \\ F_{33} \end{pmatrix} = 0. \tag{12.1.6}$$

Since (12.1.6) is homogeneous in the coefficients of $\mathcal{F}$, we can for example set $F_{33} = 1$ and use eight point correspondences $p_i \leftrightarrow p'_i$ $(i = 1, .., 8)$ to set up an 8×8

system of non-homogeneous linear equations:

$$\begin{pmatrix} u_1u_1' & u_1v_1' & u_1 & v_1u_1' & v_1v_1' & v_1 & u_1' & v_1' \\ u_2u_2' & u_2v_2' & u_2 & v_2u_2' & v_2v_2' & v_2 & u_2' & v_2' \\ u_3u_3' & u_3v_3' & u_3 & v_3u_3' & v_3v_3' & v_3 & u_3' & v_3' \\ u_4u_4' & u_4v_4' & u_4 & v_4u_4' & v_4v_4' & v_4 & u_4' & v_4' \\ u_5u_5' & u_5v_5' & u_5 & v_5u_5' & v_5v_5' & v_5 & u_5' & v_5' \\ u_6u_6' & u_6v_6' & u_6 & v_6u_6' & v_6v_6' & v_6 & u_6' & v_6' \\ u_7u_7' & u_7v_7' & u_7 & v_7u_7' & v_7v_7' & v_7 & u_7' & v_7' \\ u_8u_8' & u_8v_8' & u_8 & v_8u_8' & v_8v_8' & v_8 & u_8' & v_8' \end{pmatrix} \begin{pmatrix} F_{11} \\ F_{12} \\ F_{13} \\ F_{21} \\ F_{22} \\ F_{23} \\ F_{31} \\ F_{32} \end{pmatrix} = - \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix},$$

which is sufficient for estimating the fundamental matrix. This is the *eight-point* algorithm proposed by Longuet-Higgins [1981] in the case of calibrated cameras. It will fail when the associated 8×8 matrix is singular. As shown in [Faugeras, 1993] and the exercises, this will only happen, however, when the eight points and the two optical centers lie on a quadric surface. Fortunately, this is quite unlikely since a quadric surface is completely determined by nine points, which means that there is in general no quadric that passes through ten arbitrary points.

When $n > 8$ correspondences are available, $\mathcal{F}$ can be estimated using linear least squares by minimizing

$$\sum_{i=1}^{n} (\boldsymbol{p}_i^T \mathcal{F} \boldsymbol{p}_i')^2 \tag{12.1.7}$$

with respect to the coefficients of $\mathcal{F}$ under the constraint that the vector formed by these coefficients has unit norm.

Note that both the eight-point algorithm and its least-squares version ignore the rank-two property of fundamental matrices.[3] To enforce this constraint, Luong *et al.* [1993; 1995] have proposed to use the matrix $\mathcal{F}$ output by the eight-point algorithm as the basis for a two-step estimation process: first, use linear least squares to compute the position of the epipoles $\boldsymbol{e}$ and $\boldsymbol{e}'$ that minimize $|\mathcal{F}^T\boldsymbol{e}|^2$ and $|\mathcal{F}\boldsymbol{e}'|^2$; second, substitute the coordinates of these points in (12.1.5): this yields a linear parameterization of the fundamental matrix by the coefficients of the epipolar transformation, which can now be estimated by minimizing (12.1.7) via linear least squares.

The least-squares version of the eight-point algorithm minimizes the mean-squared *algebraic distance* associated with the epipolar constraint, i.e., the mean-squared value of $e(\boldsymbol{p}, \boldsymbol{p}') = \boldsymbol{p}^T \mathcal{F} \boldsymbol{p}'$ calculated over all point correspondences. This error function admits a geometric interpretation: in particular, we have

$$e(\boldsymbol{p}, \boldsymbol{p}') = \lambda \mathrm{d}(\boldsymbol{p}, \mathcal{F}\boldsymbol{p}') = \lambda' \mathrm{d}(\boldsymbol{p}', \mathcal{F}^T \boldsymbol{p}),$$

where $d(\boldsymbol{p}, \boldsymbol{l})$ denotes the (signed) Euclidean distance between the point $\boldsymbol{p}$ and the line $\boldsymbol{l}$, and $\mathcal{F}\boldsymbol{p}$ and $\mathcal{F}^T\boldsymbol{p}'$ are the epipolar lines associated with $\boldsymbol{p}$ and $\boldsymbol{p}'$. The

[3]The original algorithm proposed by Longuet-Higgins ignores the fact that essential matrices have rank two and two equal singular values as well.

scale factors λ and λ' are simply the norms of the vectors formed by the first two components of $\mathcal{F}\boldsymbol{p}'$ and $\mathcal{F}^T\boldsymbol{p}$, and their dependence on the pair of data points observed may bias the estimation process.

It is of course possible to get rid of the scale factors and directly minimize the mean-squared distance between the image points and the corresponding epipolar lines, i.e.,

$$\sum_{i=1}^{n}[\mathrm{d}^2(\boldsymbol{p}_i, \mathcal{F}\boldsymbol{p}_i') + \mathrm{d}^2(\boldsymbol{p}_i', \mathcal{F}^T\boldsymbol{p}_i)].$$

This is a non-linear problem, regardless of the parameterization chosen for the fundamental matrix, but the minimization can be initialized with the result of the eight-point algorithm. This method was first proposed by Luong *et al.* [1993], and it has been shown to provide results vastly superior to those obtained using the eight-point method.

Recently, Hartley [1995] has proposed a normalized eight-point algorithm and has also reported excellent results. His approach is based on the observation that the poor performance of the plain eight-point method is due, for the most part, to poor numerical conditioning. Thus Hartley has proposed to translate and scale the data so it is centered at the origin and the average distance to the origin is $\sqrt{2}$ pixel. This dramatically improves the conditioning of the linear least-squares estimation process. Accordingly, his method is divided into four steps: first, transform the image coordinates using appropriate translation and scaling operators $\mathcal{T} : \boldsymbol{p}_i \to \tilde{\boldsymbol{p}}_i$ and $\mathcal{T}' : \boldsymbol{p}_i' \to \tilde{\boldsymbol{p}}_i'$. Second, use linear least squares to compute the matrix $\tilde{\mathcal{F}}$ minimizing

$$\sum_{i=1}^{n}(\tilde{\boldsymbol{p}}_i^T\tilde{\mathcal{F}}\tilde{\boldsymbol{p}}_i')^2.$$

Third, enforce the rank-two constraint; this can be done using the two-step method of Luong *et al.* described earlier, but Hartley uses instead a technique, suggested by Tsai and Huang [1984] in the calibrated case, which constructs the *singular value decomposition* $\tilde{\mathcal{F}} = \mathcal{U}\mathcal{S}\mathcal{V}^T$ of $\tilde{\mathcal{F}}$. Here, $\mathcal{S} = \mathrm{diag}(r, s, t)$ is a diagonal 3×3 matrix with entries $r \geq s \geq t$, and $\mathcal{U}, \mathcal{V}$ are orthogonal 3×3 matrices.[4] The rank-two matrix $\bar{\mathcal{F}}$ minimizing the Frobenius norm of $\tilde{\mathcal{F}} - \bar{\mathcal{F}}$ is simply $\bar{\mathcal{F}} = \mathcal{U}\mathrm{diag}(r, s, 0)\mathcal{V}^T$ [Tsai and Huang, 1984]. Fourth, set $\mathcal{F} = \mathcal{T}^T\bar{\mathcal{F}}\mathcal{T}'$. This is the final estimate of the fundamental matrix.

Figure 12.4 shows weak calibration experiments using as input data a set of 37 point correspondences between two images of a toy house. The data points are shown in the figure as small discs, and the recovered epipolar lines are shown as short line segments. The top of the figure shows the output of the least-squares version of the plain eight-point algorithm, and the bottom part of the figure shows the results obtained using Hartley's variant of this method. As expected, the results are much better in the second case, and in fact extremely close to those obtained using the distance minimization criterion of Luong *et al.* [1993; 1995].

[4] Singular value decomposition will be discussed in detail in Chapter 14.

(a)

(b)

	Linear Least Squares	[Hartley, 1995]	[Luong *et al.*, 1993]
Av. Dist. 1	2.33 pixels	0.92 pixel	0.86 pixel
Av. Dist. 2	2.18 pixels	0.85 pixel	0.80 pixel

Figure 12.4. Weak calibration experiments using 37 point correspondences between two images of a toy house. The figure shows the epipolar lines found by (a) the least-squares version of the 8-point algorithm, and (b) the "normalized" variant of this method proposed by Hartley [1995]. Note for example the much larger error in (a) for the feature point close to the bottom of the mug. Quantitative comparisons are given in the table, where the average distances between the data points and corresponding epipolar lines are shown for both techniques as well as the non-linear distance minimization algorithm of Luong *et al.* [1993].

12.2 Three Views

Let us now go back to the calibrated case where $\boldsymbol{p} = \hat{\boldsymbol{p}}$ as we study the geometric constraints associated with three views of the same scene. Consider three perspective cameras observing the same point P, whose images are denoted by p_1, p_2 and p_3 (Figure 12.5). The optical centers O_1, O_2 and O_3 of the cameras define a *trifocal plane* T that intersects their retinas along three *trifocal lines* t_1, t_2 and t_3. Each one of these lines passes through the associated epipoles, e.g., the line t_2 associated with the second camera passes through the projections e_{12} and e_{32} of the optical centers of the two other cameras.

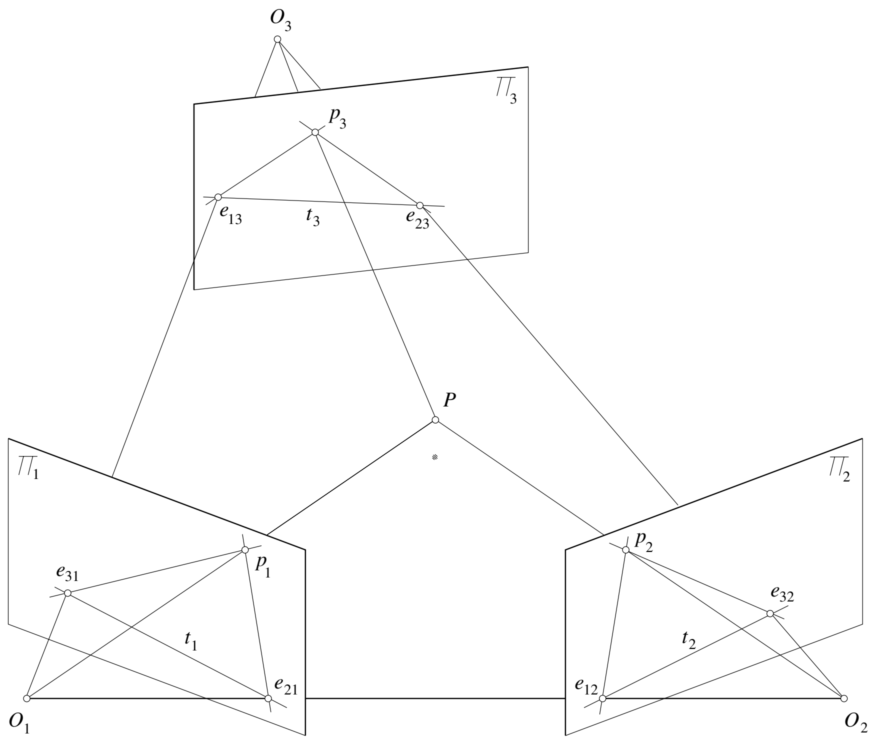


Figure 12.5. Trinocular epipolar geometry.

Each pair of cameras defines an epipolar constraint, i.e.,

$$\begin{cases} \boldsymbol{p}_1^T \mathcal{E}_{12} \boldsymbol{p}_2 = 0, \\ \boldsymbol{p}_2^T \mathcal{E}_{23} \boldsymbol{p}_3 = 0, \\ \boldsymbol{p}_3^T \mathcal{E}_{31} \boldsymbol{p}_1 = 0, \end{cases} \tag{12.2.1}$$

where $\mathcal{E}_{ij}$ denotes the essential matrix associated with the image pairs $i \leftrightarrow j$. These three constraints are not independent since we must have $\boldsymbol{e}_{31}^T \mathcal{E}_{12} \boldsymbol{e}_{32} = \boldsymbol{e}_{12}^T \mathcal{E}_{23} \boldsymbol{e}_{13} = \boldsymbol{e}_{23}^T \mathcal{E}_{31} \boldsymbol{e}_{21} = 0$ (to see why, consider for example the epipoles e_{31} and e_{32}: they are the first and second images of the optical center O_3 of the third camera, and are therefore in epipolar correspondence).

Any two of the equations in (12.2.1) are, on the other hand, independent. In particular, when the essential matrices are known, it is possible to predict the position of the point p_1 from the positions of the two corresponding points p_2 and

p_3: indeed, the first and third constraints in (12.2.1) form a system of two linear equations in the two unknown coordinates of p_1. Geometrically, p_1 is found as the intersection of the epipolar lines associated with p_2 and p_3 (Figure 12.5). Thus the trinocular epipolar geometry offers a solution to the problem of transfer mentioned in the introduction.

12.2.1 Trifocal Geometry

A second set of constraints can be obtained by considering three images of a line instead of a point: as shown in Chapter 5, the set of points that project onto an image line is the plane that contains the line and the pinhole. We can characterize this plane as follows: if $\mathcal{M}$ denotes a 3×4 projection matrix, then a point P projects onto p when $z\boldsymbol{p} = \mathcal{M}\boldsymbol{P}$, or

$$\boldsymbol{l}^T \mathcal{M}\boldsymbol{P} = 0, \tag{12.2.2}$$

where $\boldsymbol{P} = (x, y, z, 1)^T$ is the 4-vector of homogeneous coordinates of P. Equation (12.2.2) is of course the equation of the plane L that contains both the optical center O of the camera and the line l, and $\boldsymbol{L} = \mathcal{M}^T\boldsymbol{l}$ is the coordinate vector of this plane.

Two images l_1 and l_2 of the same line do not constrain the relative position and orientation of the associated cameras since the corresponding planes L_1 and L_2 always intersect (possibly at infinity). Let us now consider three images l_i, l_2 and l_3 of the same line l and denote by L_1, L_2 and L_3 the associated planes (Figure 12.6). The intersection of these planes forms a line instead of being reduced to a point in the generic case. Algebraically, this means that the system of three equations in three unknowns

$$\begin{pmatrix} \boldsymbol{L}_1^T \\ \boldsymbol{L}_2^T \\ \boldsymbol{L}_3^T \end{pmatrix} \boldsymbol{P} = \boldsymbol{0}$$

must be degenerate, or, equivalently, the rank of the 3×4 matrix

$$\mathcal{L} \stackrel{\text{def}}{=} \begin{pmatrix} \boldsymbol{l}_1^T \mathcal{M}_1 \\ \boldsymbol{l}_2^T \mathcal{M}_2 \\ \boldsymbol{l}_3^T \mathcal{M}_3 \end{pmatrix}$$

must be two, which in turn implies that the determinants of all its 3×3 minors must be zero. These determinants are clearly trilinear combinations of the coordinates vectors $\boldsymbol{l}_1$, $\boldsymbol{l}_2$ and $\boldsymbol{l}_3$. As shown below, only two of the four determinants are independent.

12.2.2 The Calibrated Case

To obtain an explicit formula for the trilinear constraints, we pick the coordinate system attached to the first camera as the world reference frame, which allows us to write the projection matrices as

$$\mathcal{M}_1 = (\,\mathrm{Id} \quad \boldsymbol{0}\,), \quad \mathcal{M}_2 = (\,\mathcal{R}_2 \quad \boldsymbol{t}_2\,) \quad \text{and} \quad \mathcal{M}_3 = (\,\mathcal{R}_3 \quad \boldsymbol{t}_3\,),$$

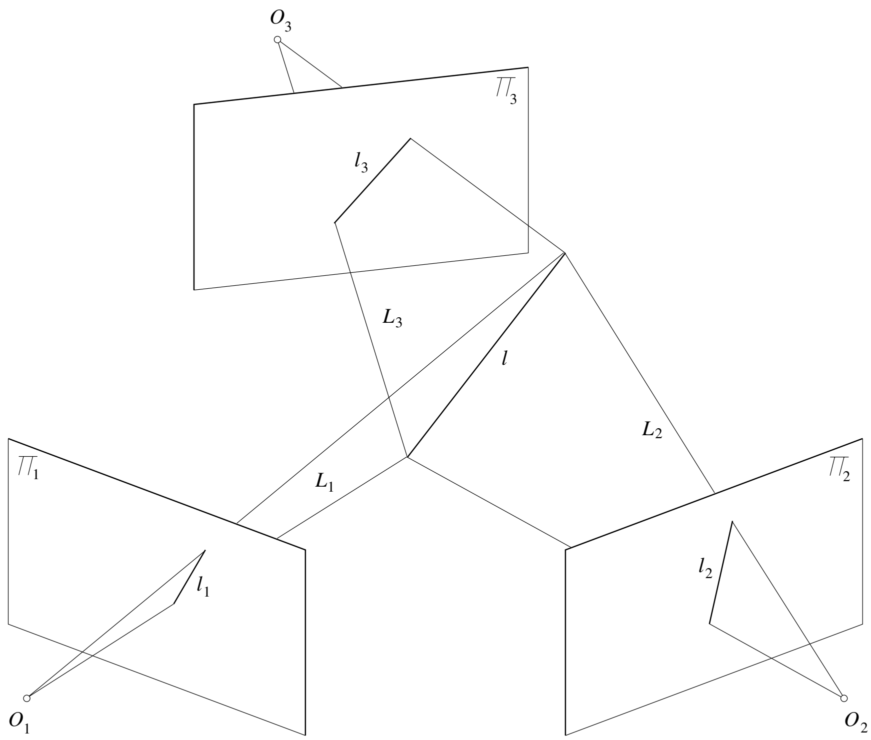


Figure 12.6. Three images of a line define it as the intersection of three planes in the same pencil.

and to rewrite $\mathcal{L}$ as

$$\mathcal{L} = \begin{pmatrix} \boldsymbol{l}_1^T & 0 \\ \boldsymbol{l}_2^T \mathcal{R}_2 & \boldsymbol{l}_2^T \boldsymbol{t}_2 \\ \boldsymbol{l}_3^T \mathcal{R}_3 & \boldsymbol{l}_3^T \boldsymbol{t}_3 \end{pmatrix}. \tag{12.2.3}$$

As shown in the exercises, three of the minor determinants can be written together as

$$\boldsymbol{l}_1 \times \begin{pmatrix} \boldsymbol{l}_2^T \mathcal{G}_1^1 \boldsymbol{l}_3 \\ \boldsymbol{l}_2^T \mathcal{G}_1^2 \boldsymbol{l}_3 \\ \boldsymbol{l}_2^T \mathcal{G}_1^3 \boldsymbol{l}_3 \end{pmatrix} = \mathbf{0}, \tag{12.2.4}$$

where

$$\mathcal{G}_1^i = \boldsymbol{t}_2 \boldsymbol{R}_3^{iT} - \boldsymbol{R}_2^i \boldsymbol{t}_3^T \quad \text{for} \quad i = 1, 2, 3, \tag{12.2.5}$$

and $\boldsymbol{R}_2^i$ and $\boldsymbol{R}_3^i$ $(i = 1, 2, 3)$ denote the columns of $\mathcal{R}_2$ and $\mathcal{R}_3$.

The fourth determinant is equal to $|\boldsymbol{l}_1 \; \mathcal{R}_2 \boldsymbol{l}_2 \; \mathcal{R}_3 \boldsymbol{l}_3|$, and it is zero when the normals to the plane L_1, L_2 and L_3 are coplanar. The corresponding equation can be written as a linear combination of the three equations in (12.2.4) (see exercises). Only two of those are linearly independent of course.

Equation (12.2.4) can finally be rewritten as

$$\boldsymbol{l}_1 \approx \begin{pmatrix} \boldsymbol{l}_2^T \mathcal{G}_1^1 \boldsymbol{l}_3 \\ \boldsymbol{l}_2^T \mathcal{G}_1^2 \boldsymbol{l}_3 \\ \boldsymbol{l}_2^T \mathcal{G}_1^3 \boldsymbol{l}_3 \end{pmatrix} \tag{12.2.6}$$

where we use $a \approx b$ to denote that a and b are equal except for an unknown scale factor.

The three 3×3 matrices $\mathcal{G}_1^i$ define the $3\times 3\times 3$ *trifocal tensor* with 27 coefficients (or 26 up to scale). (A tensor is the multi-dimensional array of coefficients associated with a multilinear form, in the same way that matrices are associated with bilinear forms.)

Since O_1 is the origin of the coordinate system in which all projection equations are expressed, the vectors $\boldsymbol{t}_2$ and $\boldsymbol{t}_3$ can be interpreted as the homogeneous image coordinates of the epipoles e_{12} and e_{13}. In particular it follows from (12.2.5) that $\boldsymbol{l}_2^T \mathcal{G}_1^i \boldsymbol{l}_3 = 0$ for any pair of matching epipolar lines l_2 and l_3.

The trifocal tensor also constrains the positions of three corresponding points. Indeed, suppose that P is a point on l. Its first image lies on l_1, so $\boldsymbol{p}_1^T \boldsymbol{l}_1 = 0$ (Figure 12.7). In particular,

$$\boldsymbol{p}_1^T \begin{pmatrix} \boldsymbol{l}_2^T \mathcal{G}_1^1 \boldsymbol{l}_3 \\ \boldsymbol{l}_2^T \mathcal{G}_1^2 \boldsymbol{l}_3 \\ \boldsymbol{l}_2^T \mathcal{G}_1^3 \boldsymbol{l}_3 \end{pmatrix} = 0. \tag{12.2.7}$$

Given three point correspondences $p_1 \leftrightarrow p_2 \leftrightarrow p_3$, we obtain four independent constraints by rewriting (12.2.7) for independent pairs of lines passing through p_2 an p_3, e.g., $\boldsymbol{l}_i' = (1, 0, -u_i)$ and $\boldsymbol{l}_i'' = (0, 1, -v_i)$ $(i = 2, 3)$. These constraints are trilinear in the coordinates of the points p_1, p_2 and p_3.

12.2.3 The Uncalibrated Case

We can still derive trilinear constraints in the image line coordinates when the intrinsic parameters of the three cameras are unknown. Since in this case $\boldsymbol{p} = \mathcal{K}\hat{\boldsymbol{p}}$, and since the image line associated with the vector $\boldsymbol{l}$ is defined by $\boldsymbol{l}^T \boldsymbol{p} = 0$, we obtain immediately $\boldsymbol{l} = \mathcal{K}^{-T}\hat{\boldsymbol{l}}$, or equivalently $\hat{\boldsymbol{l}} = \mathcal{K}^T \boldsymbol{l}$.

In particular, (12.2.3) holds when $\boldsymbol{p}_i = \hat{\boldsymbol{p}}_i$ and $\boldsymbol{l}_i = \hat{\boldsymbol{l}}_i$. In the general case we have

$$\mathcal{L} = \begin{pmatrix} \boldsymbol{l}_1^T \mathcal{K}_1 & 0 \\ \boldsymbol{l}_2^T \mathcal{K}_2 \mathcal{R}_2 & \boldsymbol{l}_2^T \mathcal{K}_2 \boldsymbol{t}_2 \\ \boldsymbol{l}_3^T \mathcal{K}_3 \mathcal{R}_3 & \boldsymbol{l}_3^T \mathcal{K}_3 \boldsymbol{t}_3 \end{pmatrix},$$

and

$$\mathrm{Rank}(\mathcal{L}) = 2 \Longleftrightarrow \mathrm{Rank}(\mathcal{L} \begin{pmatrix} \mathcal{K}_1^{-1} & 0 \\ 0 & 1 \end{pmatrix}) = \mathrm{Rank} \begin{pmatrix} \boldsymbol{l}_1^T & 0 \\ \boldsymbol{l}_2^T \mathcal{A}_2 & \boldsymbol{l}_2^T \boldsymbol{a}_2 \\ \boldsymbol{l}_3^T \mathcal{A}_3 & \boldsymbol{l}_3^T \boldsymbol{a}_3 \end{pmatrix} = 2,$$

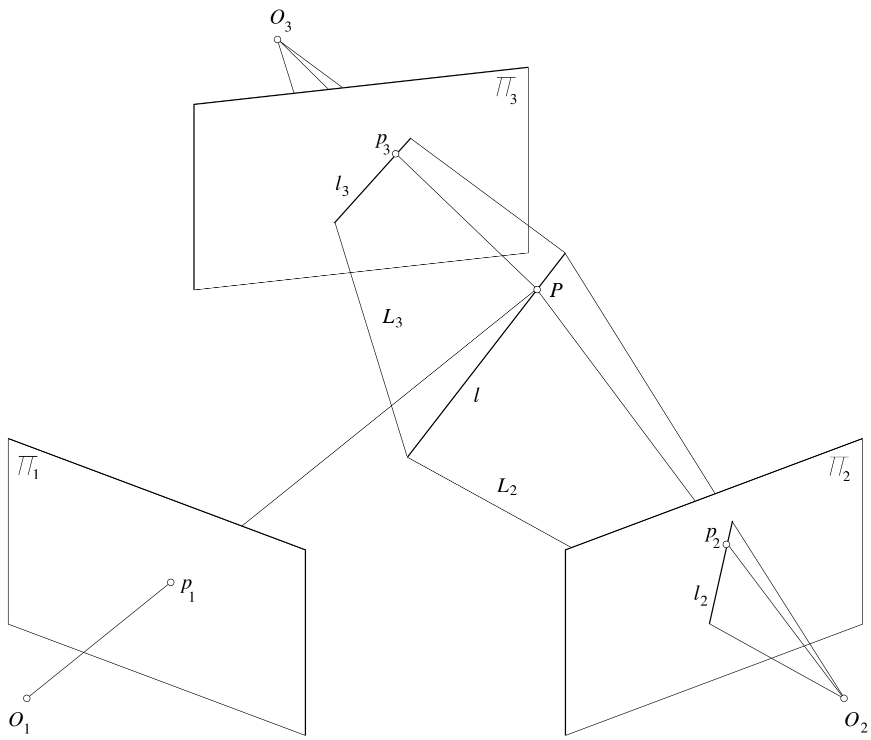


Figure 12.7. Given three images p_1, p_2 and p_3 of the same point P, and two arbitrary images l_2 and l_3 passing through the points p_2 and p_3, the ray passing through O_1 and p_1 must intersect the line where the planes L_2 and L_3 projecting onto l_2 and l_3 meet in space.

where $\mathcal{A}_i \stackrel{\text{def}}{=} \mathcal{K}_i \mathcal{R}_i \mathcal{K}_1^{-1}$ and $\boldsymbol{a}_i \stackrel{\text{def}}{=} \mathcal{K}_i \boldsymbol{t}_i$ for $i = 2, 3$. Note that the projection matrices associated with our three cameras are now $\mathcal{M}_1 = (\, \mathcal{K}_1 \quad \mathbf{0} \,)$, $\mathcal{M}_2 = (\, \mathcal{A}_2 \mathcal{K}_1 \quad \boldsymbol{a}_2 \,)$, and $\mathcal{M}_3 = (\, \mathcal{A}_3 \mathcal{K}_1 \quad \boldsymbol{a}_3 \,)$. In particular $\boldsymbol{a}_2$ and $\boldsymbol{a}_3$ can still be interpreted as the homogeneous image coordinates of the epipoles e_{12} and e_{13}, and the trilinear constraints (12.2.6) and (12.2.7) still hold when, this time,

$$\mathcal{G}_1^i = \boldsymbol{a}_2 \boldsymbol{A}_3^{iT} - \boldsymbol{A}_2^i \boldsymbol{a}_3^T,$$

where $\boldsymbol{A}_2^i$ and $\boldsymbol{A}_3^i$ $(i = 1, 2, 3)$ denote the columns of $\mathcal{A}_2$ and $\mathcal{A}_3$. As before, we will have $\boldsymbol{l}_2^T \mathcal{G}_1^i \boldsymbol{l}_3 = 0$ for any pair of matching epipolar lines l_2 and l_3.

12.2.4 Estimation of the Trifocal Tensor

We now address the problem of estimating the trifocal tensor from point and line correspondences established across triples of pictures. The equations (12.2.5) defining the tensor are linear in its coefficients and depend only on image measurements.

As in the case of weak calibration, we can therefore use linear methods to estimate these 26 parameters. Each triple of matching points provides four independent linear equations, and every triple of matching lines provides two additional linear constraints. Thus the tensor coefficients can be computed from p points and l lines granted that $2p + l \geq 13$. For example, 7 triples of points or 13 triples of lines will do the trick, as will 3 triples of points and 7 triples of lines, etc.

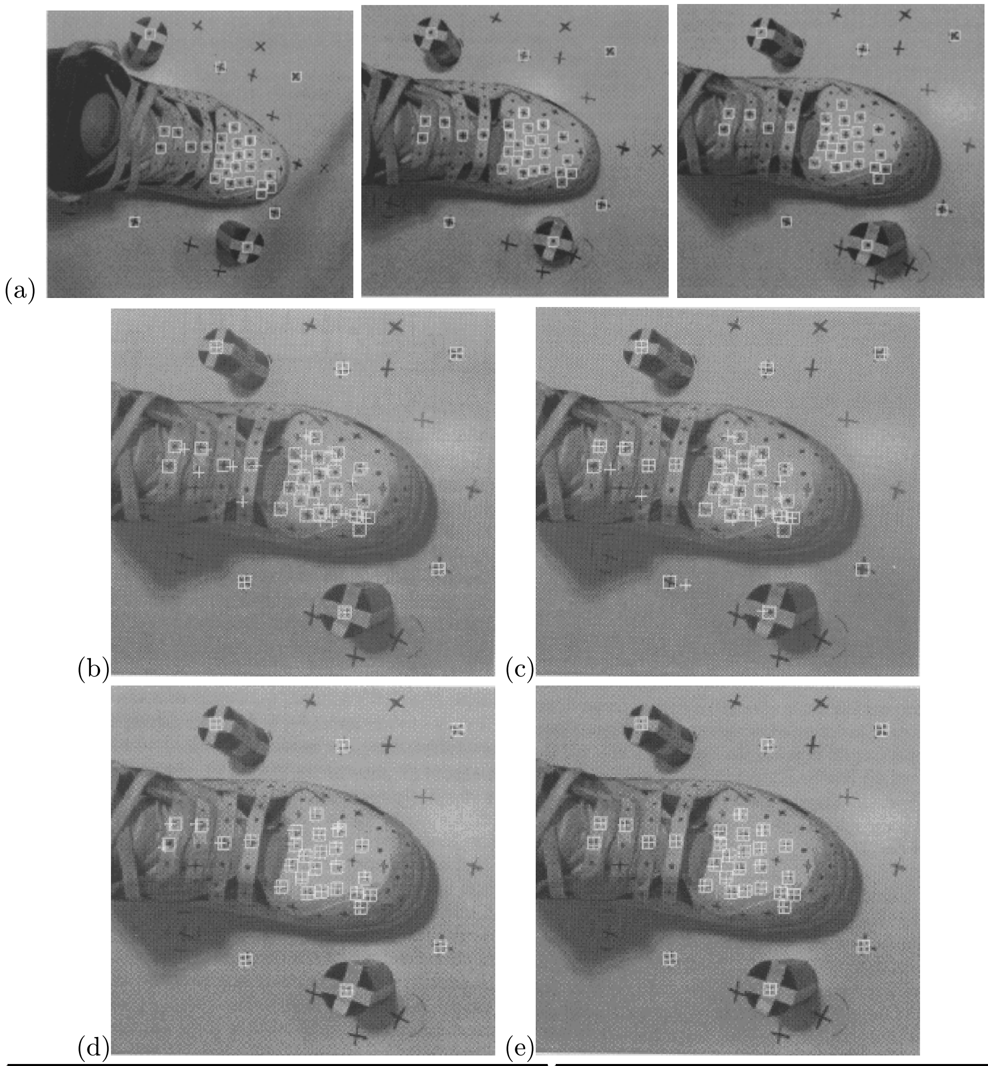
Once the tensor has been estimated, it can be used to predict the position of a point in one image from its positions in the other two. As noted before, the epipolar constraints associated with the camera pairs $1 \leftrightarrow 2$ and $1 \leftrightarrow 3$ can also be used to perform this task. Figure 12.8 shows experimental results using point correspondences found in three images of a sports shoe [Shashua, 1995]. It compares the results obtained from the fundamental matrices estimated by the method of Luong *et al.* [1993] (Figure 12.8(a)) and by a different weak-calibration technique that takes advantage of the coplanarity of correspondences lying in the ground plane supporting the shoe (see [Shashua, 1995] and Figure 12.8(b)) with the results obtained using the trifocal tensor estimated from a minimal set of seven points (Figure 12.8(c)) and a redundant set of ten correspondences (Figure 12.8(d)). In this example, the trifocal tensor clearly gives better results than the fundamental matrices.

As in the case of weak calibration, it is possible to improve the numerical stability of the tensor estimation process by normalizing the image coordinates so the data points are centered at the origin with an average distance from the origin of $\sqrt{2}$ pixel. See [Hartley, 1995] for details.

The methods outlined so far ignore the fact that the 26 parameters of the trifocal tensor are *not* independent. This should not come as a surprise: the essential matrix only has five independent coefficients (the associated rotation and translation parameters, the latter being only defined up to scale) and that the fundamental matrix only has seven. Likewise, the parameters defining the trifocal tensor satisfy a number of constraints, including the aforementioned equations $\boldsymbol{l}_2^T \mathcal{G}_1^i \boldsymbol{l}_3 = 0$ $(i = 1, 2, 3)$ satisfied by any pair of matching epipolar lines l_2 and l_3. It is also easy to show that the matrices $\mathcal{G}_1^i$ are singular, a property we will come back to in Chapter 15. Faugeras and Mourrain [1995] have shown that the coefficients of the trifocal tensor of an uncalibrated trinocular stereo rig satisfy 8 independent constraints, reducing the total number of independent parameters to 18. The method described in [Hartley, 1995] enforces these constraints *a posteriori* by recovering the epipoles e_{12} and e_{13} (or equivalently the vectors $\boldsymbol{t}_2$ and $\boldsymbol{t}_3$ in (12.2.5)) from the linearly-estimated trifocal tensor, then recovering in a linear fashion a set of tensor coefficients that satisfy the constraints.

12.3 More Views

What about four views? In this section we follow Faugeras and Mourrain [1995] and first note that we can eliminate the depth of the observed point in the projection



Transfer Using Fundamental Matrices		Transfer Using the Trifocal Tensor	
Method:	Average Distance:	Method:	Average Distance:
Ground Plane (b)	7.70 pixels	7 points (d)	0.98 pixel
[Luong *et al.*, 1993] (c)	9.58 pixels	10 points (e)	0.44 pixel

Figure 12.8. Transfer experiments: (a) input images; (b)-(c) transfer using the fundamental matrix, estimated in (a) using correspondences on the ground floor and in (b) using the non-linear method of [Luong *et al.*, 1993]; (d)-(e) transfer using the trifocal tensor estimated in (d) from seven points, and in (e) using least squares from ten points. Reprinted from [Shashua, 1995], Figures 2–4. Quantitative comparisons are given in the table, where the average distances between the data points and the transfered ones are shown for each method. The input features are indicated by white squares and the reprojected ones are are indicated by white crosses.

equation by writing

$$z\boldsymbol{p} = \mathcal{M}\boldsymbol{P} \Longleftrightarrow \boldsymbol{p} \times (\mathcal{M}\boldsymbol{P}) = ([\boldsymbol{p}_\times]\mathcal{M})\boldsymbol{P} = 0. \tag{12.3.1}$$

Of course, only two of the scalar equations associated with this vector equation are independent. Choosing (for example) the first and second of these equations allows us to rewrite (12.3.1) as

$$\begin{pmatrix} u\mathcal{M}^3 - \mathcal{M}^1 \\ v\mathcal{M}^3 - \mathcal{M}^2 \end{pmatrix} \boldsymbol{P} = 0, \tag{12.3.2}$$

where $\mathcal{M}^i$ denotes row number i of the matrix $\mathcal{M}$.

Suppose now that we have four views, with associated projection matrices $\mathcal{M}_i$ ($i = 1, 2, 3, 4$). Writing (12.3.2) for each one of these yields

$$\mathcal{Q}\boldsymbol{P} = 0, \quad \text{where} \quad \mathcal{Q} \stackrel{\text{def}}{=} \begin{pmatrix} u_1\mathcal{M}_1^3 - \mathcal{M}_1^1 \\ v_1\mathcal{M}_1^3 - \mathcal{M}_1^2 \\ u_2\mathcal{M}_2^3 - \mathcal{M}_2^1 \\ v_2\mathcal{M}_2^3 - \mathcal{M}_2^2 \\ u_3\mathcal{M}_3^3 - \mathcal{M}_3^1 \\ v_3\mathcal{M}_3^3 - \mathcal{M}_3^2 \\ u_4\mathcal{M}_4^3 - \mathcal{M}_4^1 \\ v_4\mathcal{M}_4^3 - \mathcal{M}_4^2 \end{pmatrix}. \tag{12.3.3}$$

Equation (12.3.3) is a system of eight homogeneous equations in four unknowns that admits a non-trivial solution. It follows that the rank of the corresponding 8×4 matrix $\mathcal{Q}$ is at most 3, or, equivalently, all its 4×4 minors must have zero determinants. Geometrically, each pair of equations in (12.3.3) represents the ray R_i ($i = 1, 2, 3, 4$) associated with the image point p_i, and $\mathcal{Q}$ must have rank 3 for these rays to intersect at a point P (Figure 12.9).

The matrix $\mathcal{Q}$ has three kinds of 4×4 minors:

1. Those that involve two rows from one projection matrix, and two rows from another one. The equations associated with the six minors of this type include, for example,[5]

$$\text{Det} \begin{pmatrix} u_2\mathcal{M}_1^3 - \mathcal{M}_1^1 \\ v_2\mathcal{M}_1^3 - \mathcal{M}_1^2 \\ u_3\mathcal{M}_2^3 - \mathcal{M}_2^1 \\ v_3\mathcal{M}_2^3 - \mathcal{M}_2^2 \end{pmatrix} = 0. \tag{12.3.4}$$

   These determinants yield bilinear constraints on the position of the associated image points. It is easy to show (see exercises) that the corresponding equations reduce to the epipolar constraints (12.1.2) when we take $\mathcal{M}_1 = (\,\text{Id} \quad 0\,)$ and $\mathcal{M}_2 = (\,\mathcal{R}^T \quad -\mathcal{R}^T\boldsymbol{t}\,)$.

[5] General formulas can be given as well by using for example (u^1, u^2) instead of (u, v) and playing around with indices and tensorial notation. We will abstain from this worthy exercise here.

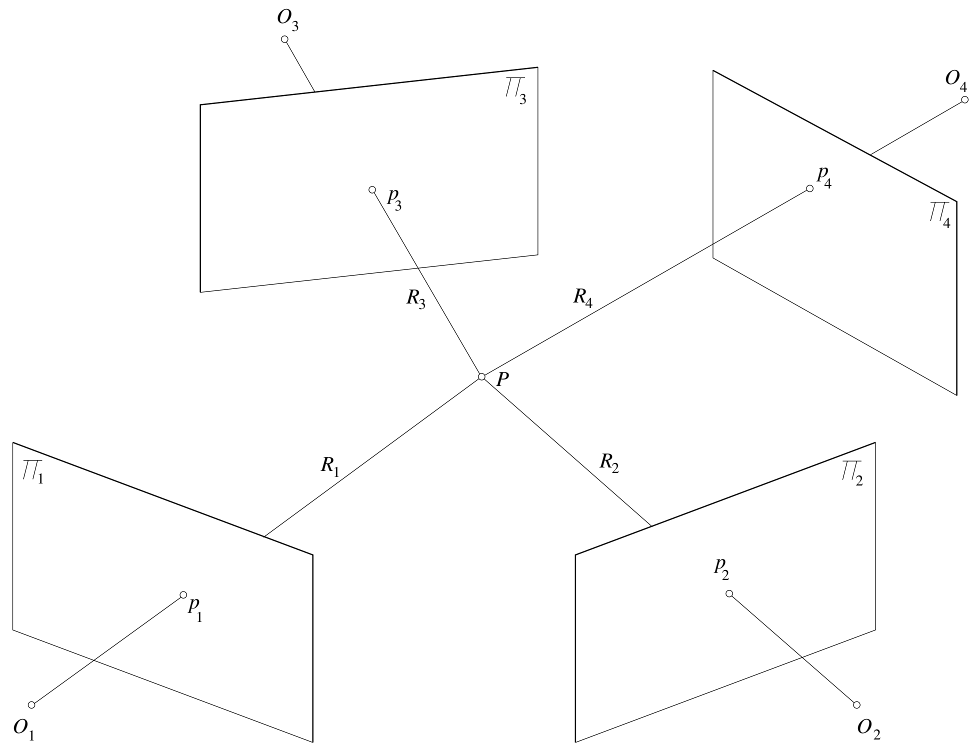


Figure 12.9. Four images p_1, p_2, p_3 and p_4 of the same point P define this point as the intersection of the corresponding rays R_i $(i = 1, 2, 3, 4)$.

2. The second type of minors involves two rows from one projection matrix, and one row from each of two other matrices. There are 48 of those, and the associated equations include, for example,

$$\text{Det}\begin{pmatrix} u_1\mathcal{M}_1^3 - \mathcal{M}_1^1 \\ v_1\mathcal{M}_1^3 - \mathcal{M}_1^2 \\ u_2\mathcal{M}_2^3 - \mathcal{M}_2^1 \\ v_3\mathcal{M}_3^3 - \mathcal{M}_3^2 \end{pmatrix} = 0. \tag{12.3.5}$$

These minors yield trilinear constraints on the corresponding image positions. It is easy to show (see exercises) that the corresponding equations reduce to the trifocal constraints (12.2.7) introduced in the previous section when we take $\mathcal{M}_1 = (\,\text{Id} \quad 0\,)$. In particular, they can be expressed in terms of the matrices $\mathcal{G}_1^i$ $(i = 1, 2, 3)$. Note that this completes the geometric interpretation of the trifocal constraints, that express here the fact that the rays associated with three images of the same point must intersect in space.

3. The last type of determinant involves one row of each matrix. The equations associated with the 16 minors of this form include, for example,

$$\operatorname{Det}\begin{pmatrix} v_1\mathcal{M}_1^3 - \mathcal{M}_1^2 \\ u_2\mathcal{M}_2^3 - \mathcal{M}_2^1 \\ v_3\mathcal{M}_3^3 - \mathcal{M}_3^2 \\ v_4\mathcal{M}_4^3 - \mathcal{M}_4^2 \end{pmatrix} = 0. \tag{12.3.6}$$

   These equations yield *quadrilinear* constraints on the position of the points $\boldsymbol{p}_i$ $(i = 1, 2, 3, 4)$. Geometrically, each row of the matrix $\mathcal{Q}$ is associated with an image line or equivalently with a plane passing through the optical center of the corresponding camera. Thus each quadrilinearity expresses the fact that the four associated planes intersect in a point (instead of not intersecting at all in the generic case).

Let us focus from now on the the quadrilinear equations. Developing determinants such as (12.3.6) with respect to the image coordinates reveals immediately that the coefficients of the quadrilinear constraints can be written as

$$\varepsilon_{ijkl}\operatorname{Det}\begin{pmatrix} \mathcal{M}_1^i \\ \mathcal{M}_2^j \\ \mathcal{M}_3^k \\ \mathcal{M}_4^l \end{pmatrix}, \tag{12.3.7}$$

where $\varepsilon_{ijkl} = \mp 1$ and i, j, k and l are indices between 1 and 4 (see exercises). These coefficients determine the *quadrifocal tensor* [Triggs, 1995].

Like its trifocal cousin, this tensor can be interpreted geometrically using both points and lines. In particular, consider four pictures p_i $(i = 1, 2, 3, 4)$ of a point P and four arbitrary image lines l_i passing through these points. The four planes L_i $(i = 1, 2, 3, 4)$ formed by the preimages of the lines must intersect in P, which implies in turn that the 4×4 matrix

$$\mathcal{L} \stackrel{\text{def}}{=} \begin{pmatrix} \boldsymbol{l}_1^T\mathcal{M}_1 \\ \boldsymbol{l}_2^T\mathcal{M}_2 \\ \boldsymbol{l}_3^T\mathcal{M}_3 \\ \boldsymbol{l}_4^T\mathcal{M}_4 \end{pmatrix}$$

must have rank 3, and, in particular, that its determinant must be zero. This obviously provides a quadrilinear constraint on the coefficients of the four lines l_i $(i = 1, 2, 3, 4)$. In addition, since each row $\boldsymbol{L}_i^T = \boldsymbol{l}_i^T\mathcal{M}_i$ of $\mathcal{L}$ is a linear combination of the rows of the associated matrix $\mathcal{M}_i$, the coefficients of the quadrilinearities obtained by developing $\operatorname{Det}(\mathcal{L})$ with respect to the coordinates of the lines l_i are simply the coefficients of the quadrifocal tensor as defined by (12.3.7).

Finally, note since $\operatorname{Det}(\mathcal{L})$ is linear in the coordinates of $\boldsymbol{l}_1$, the vanishing of this determinant can be written as $\boldsymbol{l}_1 \cdot \boldsymbol{q}(\boldsymbol{l}_2, \boldsymbol{l}_3, \boldsymbol{l}_4) = 0$, where $\boldsymbol{q}$ is a (trilinear) function of the coordinates of the lines l_i $(i = 2, 3, 4)$. Since this relationship holds for any line

l_1 passing through p_1 it follows that $\boldsymbol{p}_1 \approx \boldsymbol{q}(\boldsymbol{l}_2, \boldsymbol{l}_3, \boldsymbol{l}_4)$. Geometrically, this means that the ray passing through O_1 and p_1 must also pass through the intersection of the planes formed by the preimages of l_2, l_3 and l_4 (Figure 12.10). Algebraically, this means that, given the quadrifocal tensor and arbitrary lines passing through three images of a point, we can predict the position of this point in a fourth image. This provides yet another method for transfer.

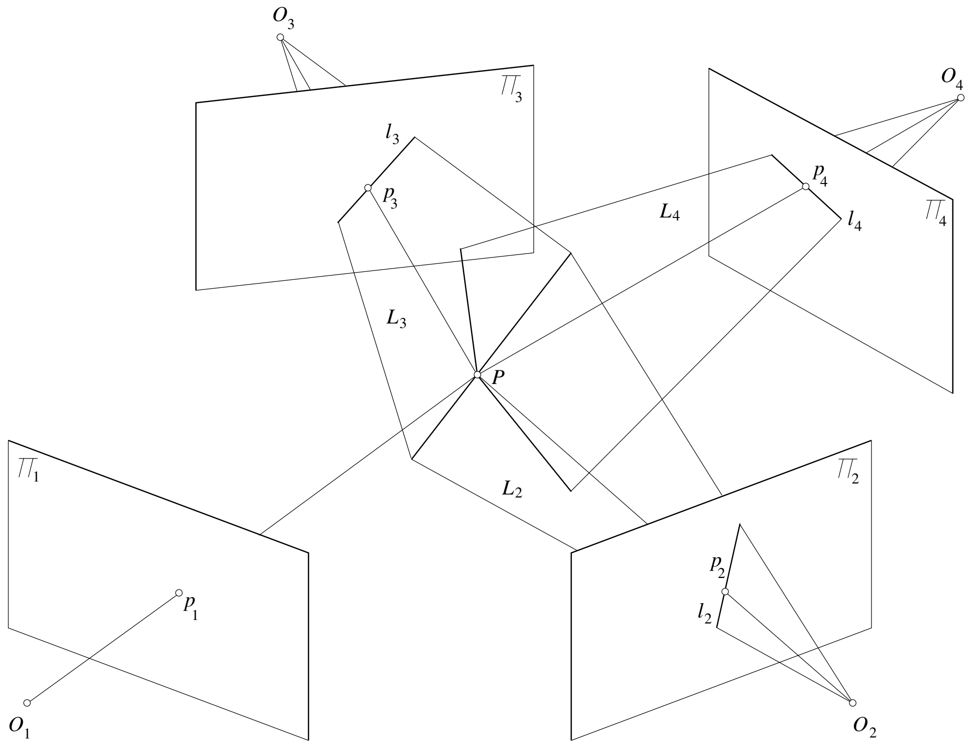


Figure 12.10. Given four images p_1, p_2, p_3 and p_4 of some point P and three arbitrary image lines l_2, l_3 and l_4 passing through the points p_2, p_3 and p_4, the ray passing through O_1 and p_1 must also pass through the point where the three planes L_2, L_3 and L_4 formed by the preimages of these lines intersect.

Note that the quadrifocal constraints are valid in both the calibrated and uncalibrated cases since we have made no assumption on the form of the matrices $\mathcal{M}_i$. The quadrifocal tensor is defined by 81 coefficients (or 80 up to scale), but these coefficients satisfy 51 independent constraints, reducing the total number of independent parameters to 29 [Heyden, 1998; Hartley, 1998]. It can also be shown that, although each quadruple of images of the same point yields 16 independent constraints like (12.3.6) on the 80 tensor coefficients, there exists a linear dependency between the 32 equations associated with each pair of points [Heyden, 1998].

Thus six point correspondences are necessary to estimate the quadrifocal tensor in a linear fashion. Algorithms for performing this task and enforcing the 51 constraints associated with actual quadrifocal tensors can be found in [Hartley, 1998].

Finally, Faugeras and Mourrain [1995] have shown that the quadrilinear tensor is algebraically dependent on the associated essential/fundamental matrices and trifocal tensor, and thus does not add independent new constraints. Likewise, it can be shown that additional views do not add independent constraints either.

12.4 Notes

The essential matrix as an algebraic form of the epipolar constraint was discovered by Longuet-Higgins [1981], and its properties have been elucidated by Huang and Faugeras [1989]. The fundamental matrix was introduced by Luong and Faugeras [1992; 1995]. We will come back to the properties of the fundamental matrix and of the epipolar transformation in Chapter 15, when we adress the problem of recovering the structure of a scene and the motion of a camera from a sequence of perspective images.

The trilinear constraints associated with three views of a line were introduced independently by Spektakis and Aloimonos [1990] and Weng, Huang and Ahuja [1992] in the context of motion analysis for internally calibrated cameras. They were extended by Shashua [1995] and Hartley [1997] to the uncalibrated case. The quadrifocal tensor was introduced by Triggs [1995]. Geometric studies can by found in Faugeras and Mourrain [1995], Faugeras and Papadopoulo [1997] and Heyden [1998].

We mentioned in the introduction that photogrammetry is concerned with the extraction of quantitative information from multiple pictures. In this context, binocular and trinocular geometric constraints are regarded as the source of *condition equations* that determine the intrinsic and extrinsic parameters (called *interior* and *exterior orientation* parameters in photogrammetry) of a stereo pair or triple. In particular, the Longuet-Higgins relation appears, in a slightly disguised form, as the *coplanarity condition equation*, and trinocular constraints yield *scale-restraint condition equations* that take calibration and image measurement errors into account [Thompson *et al.*, 1966, Chapter X]: in this case, the rays associated with three images of the same point are not guaranteed to intersect anymore (Figure 12.11).

The setup is as follows: if the rays R_1 and R_i ($i = 2, 3$) associated with the image points p_1 and p_i do not intersect, the minimum distance between them is reached at the points P_1 and P_i such that the line joining these points is perpendicular to both R_1 and R_i. Algebraically, this can be written as

$$\overrightarrow{O_1P_1} = z_1^i \overrightarrow{O_1p_1} = \overrightarrow{O_1O_i} + z_i \overrightarrow{O_ip_i} + \lambda_i(\overrightarrow{O_1p_1} \times \overrightarrow{O_ip_i}) \quad \text{for} \quad i = 2, 3. \tag{12.4.1}$$

Assuming that the cameras are internally calibrated so the projection matrices associated with the second and third cameras are $(\mathcal{R}_2^T \quad -\mathcal{R}_2^T \boldsymbol{t}_2)$ and $(\mathcal{R}_3^T \quad -\mathcal{R}_3^T \boldsymbol{t}_3)$, (12.4.1) can be rewritten in the coordinate system attached to the first camera as

$$z_1^i \boldsymbol{p}_1 = \boldsymbol{t}_i + z_i \mathcal{R}_i \boldsymbol{p}_i + \lambda_i(\boldsymbol{p}_1 \times \mathcal{R}_i \boldsymbol{p}_i) \quad \text{for} \quad i = 2, 3. \tag{12.4.2}$$

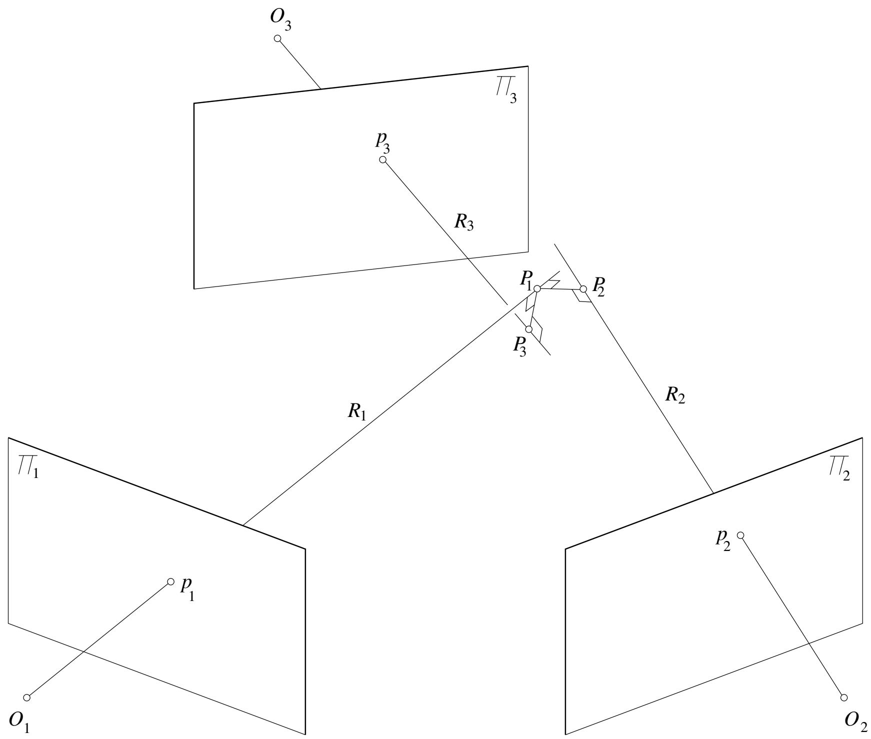


Figure 12.11. Trinocular constraints in the presence of calibration or measurement errors: the rays R_1, R_2 and R_3 may not intersect.

Note that a similar equation could be written as well for completely uncalibrated cameras by including terms depending on the (unknown) intrinsic parameters. In either case, (12.4.2) can be used to calculate the unknowns z_i, λ_i and z_1^i in terms of $\boldsymbol{p}_1$, $\boldsymbol{p}_i$, and the projection matrices associated with the cameras (see exercises). The scale-restraint condition is then written as $z_1^2 = z_1^3$. Although it is more complex than the trifocal constraint (in particular, it is not trilinear in the coordinates of the points p_1, p_2 and p_3), this condition does not involve the coordinates of the observed point, and it can be used (in principle) to estimate the trifocal geometry directly from image data. A potential advantage is that the error function $z_1^2 - z_1^3$ has a clear geometric meaning: it is the difference between the estimates of the depth of P obtained using the pairs of cameras $1 \leftrightarrow 2$ and $1 \leftrightarrow 3$. It would be interesting to further investigate the relationship between the trifocal tensor and the scale-constraint condition, as well as its practical application to the estimation of the trifocal geometry.

12.5 Assignments

Exercises

1. Show that one of the singular values of an essential matrix is 0 and the other two are equal. (Huang and Faugeras [1989] have shown that the converse is also true, i.e., any 3×3 matrix with one singular value equal to 0 and the other two equal to each other is an essential matrix.)

   Hint: the singular values of $\mathcal{E}$ are the eigenvalues of $\mathcal{E}\mathcal{E}^T$.

   Solution: We have $\mathcal{E} = [\boldsymbol{t}_\times]\mathcal{R}$, thus $\mathcal{E}\mathcal{E}^T = [\boldsymbol{t}_\times][\boldsymbol{t}_\times]^T = [\boldsymbol{t}_\times]^T[\boldsymbol{t}_\times]$. If $\boldsymbol{a}$ is an eigenvector of $\mathcal{E}\mathcal{E}^T$ associated with the eigenvalue λ then, for any vector $\boldsymbol{b}$

   $$\lambda \boldsymbol{b} \cdot \boldsymbol{a} = \boldsymbol{b}^T([\boldsymbol{t}_\times]^T[\boldsymbol{t}_\times]\boldsymbol{a}) = (\boldsymbol{t} \times \boldsymbol{b}) \cdot (\boldsymbol{t} \times \boldsymbol{a}).$$

   Choosing $\boldsymbol{a} = \boldsymbol{b} = \boldsymbol{t}$ shows that $\lambda = 0$ is an eigenvalue of $\mathcal{E}\mathcal{E}^T$. Choosing $\boldsymbol{b} = \boldsymbol{t}$ shows that if $\lambda \neq 0$ then $\boldsymbol{a}$ is orthogonal to $\boldsymbol{t}$. But then choosing $\boldsymbol{a} = \boldsymbol{b}$ shows that

   $$\lambda|\boldsymbol{a}|^2 = |\boldsymbol{t} \times \boldsymbol{a}|^2 = |\boldsymbol{t}|^2|\boldsymbol{a}|^2.$$

   It follows that all non-zero singular values of $\mathcal{E}$ must be equal. Note that the singular values of $\mathcal{E}$ cannot all be zero since this matrix has rank 2.

2. The infinitesimal epipolar constraint (12.1.3) was derived by assuming that the observed scene was static and the camera was moving. Show that when the camera is fixed and the scene is moving with translational velocity $\boldsymbol{v}$ and rotational velocity $\boldsymbol{\omega}$, the epipolar constraint can be rewritten as

   $$\boldsymbol{p}^T([\boldsymbol{v}_\times][\boldsymbol{\omega}_\times])\boldsymbol{p} + (\boldsymbol{p} \times \dot{\boldsymbol{p}}) \cdot \boldsymbol{v} = 0.$$

   Note that this equation is now the sum of the two terms appearing in (12.1.3) instead of their difference.

   Hint: If $\mathcal{R}$ and $\boldsymbol{t}$ denote the rotation matrix and translation vectors appearing in the definition of the essential matrix for a moving camera, show that the object displacement that yields the same motion field for a static camera is given by the rotation matrix $\mathcal{R}^T$ and the translation vector $-\mathcal{R}^T\boldsymbol{t}$.

3. Show that when the 8×8 matrix associated with the eight-point algorithm is singular, the eight points and the two optical centers lie on a quadric surface [Faugeras, 1993].

   Hint: Use the fact that when a matrix is singular, there exists some non-trivial linear combination of its columns that is equal to zero. Also take advantage of the fact that the matrices representing the two projections in the coordinate system of the first camera are in this case $(\,\mathrm{Id} \quad 0\,)$ and $(\,\mathcal{R}^T, -\mathcal{R}^T\boldsymbol{t}\,)$.

4. Show that three of the determinants of the 3×3 minors of
$$\mathcal{L} = \begin{pmatrix} \boldsymbol{l}_1^T & 0 \\ \boldsymbol{l}_2^T \mathcal{R}_2 & \boldsymbol{l}_2^T \boldsymbol{t}_2 \\ \boldsymbol{l}_3^T \mathcal{R}_3 & \boldsymbol{l}_3^T \boldsymbol{t}_3 \end{pmatrix}.$$
can be written as
$$\boldsymbol{l}_1 \times \begin{pmatrix} \boldsymbol{l}_2^T \mathcal{G}_1^1 \boldsymbol{l}_3 \\ \boldsymbol{l}_2^T \mathcal{G}_1^2 \boldsymbol{l}_3 \\ \boldsymbol{l}_2^T \mathcal{G}_1^3 \boldsymbol{l}_3 \end{pmatrix} = \boldsymbol{0}.$$
Show that the fourth determinant can be written as a linear combination of these.

5. Show that (12.3.4) reduces to (12.1.2) when $\mathcal{M}_1 = (\,\mathrm{Id} \quad 0\,)$ and $\mathcal{M}_2 = (\,\mathcal{R}^T \quad -\mathcal{R}^T \boldsymbol{t}\,)$.

6. Show that (12.3.5) reduces to (12.2.7) when $\mathcal{M}_1 = (\,\mathrm{Id} \quad 0\,)$.

7. Develop (12.3.6) with respect to the image coordinates and verify that the coefficients can indeed be written in the form (12.3.7).

8. Use (12.4.2) to calculate the unknowns z_i, λ_i and z_1^i in terms of $\boldsymbol{p}_1$, $\boldsymbol{p}_i$, $\mathcal{R}_i$ and $\boldsymbol{t}_i$ ($i = 2, 3$). Show that the value of λ_i is directly related to the epipolar constraint and characterize the degree of the dependency of $z_1^2 - z_1^3$ on the data points.

Programming Assignments

Note: the assignments below require routines for solving square and overdetermined linear systems. An extensive set of such routines is available in MATLAB as well as in public-domain libraries such as LINPACK and LAPACK that can be downloaded from the Netlib repository (`http://www.netlib.org/`). Data for these assignments will be available in the CD companion to this book.

1. Implement the 8-point algorithm for weak calibration from binocular point correspondences.

2. Implement the linear least-squares version of that algorithm with and without Hartley's pre-conditioning step.

3. Implement an algorithm for estimating the trifocal tensor from point correspondences.

4. Implement an algorithm for estimating the trifocal tensor from line correspondences.