15

The Trifocal Tensor

The trifocal tensor plays an analogous role in three views to that played by the fundamental matrix in two. It encapsulates all the (projective) geometric relations between three views that are independent of scene structure.

We begin this chapter with a simple introduction to the main geometric and algebraic properties of the trifocal tensor. A formal development of the trifocal tensor and its properties involves the use of tensor notation. To start, however, it is convenient to use standard vector and matrix notation, thus obtaining some geometric insight into the trifocal tensor without the additional burden of dealing with a (possibly) unfamiliar notation. The use of tensor notation will therefore be deferred until section 15.2.

The three principal geometric properties of the tensor are introduced in section 15.1. These are the homography between two of the views induced by a plane back-projected from a line in the other view; the relations between image correspondences for points and lines which arise from incidence relations in 3-space; and the retrieval of the fundamental and camera matrices from the tensor.

The tensor may be used to transfer points from a correspondence in two views to the corresponding point in a third view. The tensor also applies to lines, and the image of a line in one view may be computed from its corresponding images in two other views. Transfer is described in section 15.3.

The tensor only depends on the motion between views and the internal parameters of the cameras and is defined uniquely by the camera matrices of the views. However, it can be computed from image correspondences alone without requiring knowledge of the motion or calibration. This computation is described in chapter 16.

15.1 The geometric basis for the trifocal tensor

There are several ways that the trifocal tensor may be approached, but in this section the starting point is taken to be the incidence relationship of three corresponding lines.

Incidence relations for lines. Suppose a line in 3-space is imaged in three views, as in figure 15.1, what constraints are there on the corresponding image lines? The planes back-projected from the lines in each view must all meet in a single line in space, the 3D line that projects to the matched lines in the three images. Since in general three arbitrary planes in space do not meet in a single line, this geometric *incidence* condition

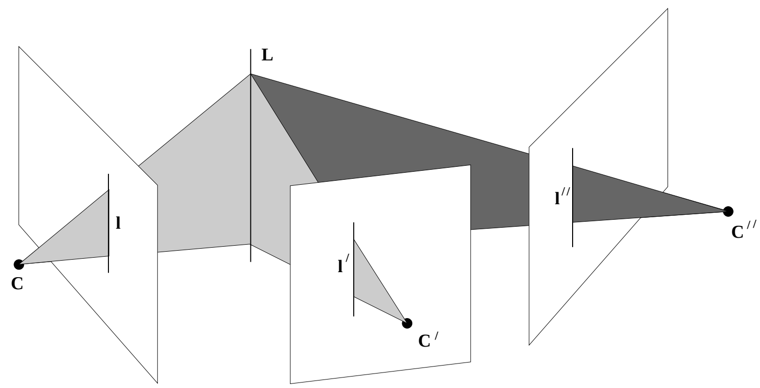


Fig. 15.1. *A line $\mathbf{L}$ in 3-space is imaged as the corresponding triplet $\mathbf{l} \leftrightarrow \mathbf{l}' \leftrightarrow \mathbf{l}''$ in three views indicated by their centres, $\mathbf{C}, \mathbf{C}', \mathbf{C}''$, and image planes. Conversely, corresponding lines back-projected from the first, second and third images all intersect in a single 3D line in space.*

provides a genuine constraint on sets of corresponding lines. We will now translate this geometric constraint into an algebraic constraint on the three lines.

We denote a set of corresponding lines as $\mathbf{l}_i \leftrightarrow \mathbf{l}'_i \leftrightarrow \mathbf{l}''_i$. Let the camera matrices for the three views be $\mathtt{P} = [\mathtt{I} \mid \mathbf{0}]$, as usual, and $\mathtt{P}' = [\mathtt{A} \mid \mathbf{a}_4]$, $\mathtt{P}'' = [\mathtt{B} \mid \mathbf{b}_4]$, where $\mathtt{A}$ and $\mathtt{B}$ are 3×3 matrices, and the vectors $\mathbf{a}_i$ and $\mathbf{b}_i$ are the i-th columns of the respective camera matrices for $i = 1, \ldots, 4$.

- $\mathbf{a}_4$ and $\mathbf{b}_4$ are the epipoles in views two and three respectively, arising from the first camera. These epipoles will be denoted by $\mathbf{e}'$ and $\mathbf{e}''$ throughout this chapter, with $\mathbf{e}' = \mathtt{P}'\mathbf{C}$, $\mathbf{e}'' = \mathtt{P}''\mathbf{C}$, where $\mathbf{C}$ is the first camera centre. (For the most part we will not be concerned with the epipoles between the second and third views).
- $\mathtt{A}$ and $\mathtt{B}$ are the infinite homographies from the first to the second and third cameras respectively.

As has been seen in chapter 9, any set of three cameras is equivalent to a set with $\mathtt{P} = [\mathtt{I} \mid \mathbf{0}]$ under projective transformations of space. In this chapter we will be concerned with properties (such as image coordinates and 3D incidence relations) that are invariant under 3D projective transforms, so we are free to choose the cameras in this form.

Now, each image line back-projects to a plane, as shown in figure 15.1. From result 8.2(*p*197) these three planes are

$$\boldsymbol{\pi} = \mathtt{P}^{\mathsf{T}}\mathbf{l} = \begin{pmatrix} \mathbf{l} \\ 0 \end{pmatrix} \quad \boldsymbol{\pi}' = \mathtt{P}'^{\mathsf{T}}\mathbf{l}' = \begin{pmatrix} \mathtt{A}^{\mathsf{T}}\mathbf{l}' \\ \mathbf{a}_4^{\mathsf{T}}\mathbf{l}' \end{pmatrix} \quad \boldsymbol{\pi}'' = \mathtt{P}''^{\mathsf{T}}\mathbf{l}'' = \begin{pmatrix} \mathtt{B}^{\mathsf{T}}\mathbf{l}'' \\ \mathbf{b}_4^{\mathsf{T}}\mathbf{l}'' \end{pmatrix}.$$

Since the three image lines are derived from a single line in space, it follows that these three planes are not independent but must meet in this common line in 3-space. This intersection constraint can be expressed algebraically by the requirement that the 4×3 matrix $\mathtt{M} = [\boldsymbol{\pi}, \boldsymbol{\pi}', \boldsymbol{\pi}'']$ has rank 2. This may be seen as follows. Points on the line of intersection may be represented as $\mathbf{X} = \alpha\mathbf{X}_1 + \beta\mathbf{X}_2$, with $\mathbf{X}_1$ and $\mathbf{X}_2$ linearly independent. Such points lie on all three planes and so $\boldsymbol{\pi}^{\mathsf{T}}\mathbf{X} = \boldsymbol{\pi}'^{\mathsf{T}}\mathbf{X} = \boldsymbol{\pi}''^{\mathsf{T}}\mathbf{X} = 0$. It follows that $\mathtt{M}^\mathsf{T}\mathbf{X} = \mathbf{0}$. Consequently $\mathtt{M}$ has a 2-dimensional null-space since $\mathtt{M}^\mathsf{T}\mathbf{X}_1 = \mathbf{0}$ and $\mathtt{M}^\mathsf{T}\mathbf{X}_2 = \mathbf{0}$.

This intersection constraint induces a relation amongst the image lines $\mathbf{l}, \mathbf{l}', \mathbf{l}''$. Since the rank of $\mathtt{M}$ is 2, there is a linear dependence between its columns $\mathbf{m}_i$. Denoting

$$\mathtt{M} = [\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3] = \begin{bmatrix} \mathbf{l} & \mathtt{A}^\mathsf{T}\mathbf{l}' & \mathtt{B}^\mathsf{T}\mathbf{l}'' \\ 0 & \mathbf{a}_4^\mathsf{T}\mathbf{l}' & \mathbf{b}_4^\mathsf{T}\mathbf{l}'' \end{bmatrix}$$

the linear relation may be written $\mathbf{m}_1 = \alpha\mathbf{m}_2 + \beta\mathbf{m}_3$. Then noting that the bottom left hand element of $\mathtt{M}$ is zero, it follows that $\alpha = k(\mathbf{b}_4^\mathsf{T}\mathbf{l}'')$ and $\beta = -k(\mathbf{a}_4^\mathsf{T}\mathbf{l}')$ for some scalar k. Applying this to the top 3-vectors of each column shows that (up to a homogeneous scale factor)

$$\mathbf{l} = (\mathbf{b}_4^\mathsf{T}\mathbf{l}'')\mathtt{A}^\mathsf{T}\mathbf{l}' - (\mathbf{a}_4^\mathsf{T}\mathbf{l}')\mathtt{B}^\mathsf{T}\mathbf{l}'' = (\mathbf{l}''^\mathsf{T}\mathbf{b}_4)\mathtt{A}^\mathsf{T}\mathbf{l}' - (\mathbf{l}'^\mathsf{T}\mathbf{a}_4)\mathtt{B}^\mathsf{T}\mathbf{l}''.$$

The i-th coordinate l_i of $\mathbf{l}$ may therefore be written as

$$l_i = \mathbf{l}''^\mathsf{T}(\mathbf{b}_4\mathbf{a}_i^\mathsf{T})\mathbf{l}' - \mathbf{l}'^\mathsf{T}(\mathbf{a}_4\mathbf{b}_i^\mathsf{T})\mathbf{l}'' = \mathbf{l}'^\mathsf{T}(\mathbf{a}_i\mathbf{b}_4^\mathsf{T})\mathbf{l}'' - \mathbf{l}'^\mathsf{T}(\mathbf{a}_4\mathbf{b}_i^\mathsf{T})\mathbf{l}''$$

and introducing the notation

$$\mathtt{T}_i = \mathbf{a}_i\mathbf{b}_4^\mathsf{T} - \mathbf{a}_4\mathbf{b}_i^\mathsf{T} \tag{15.1}$$

the incidence relation can be written

$$l_i = \mathbf{l}'^\mathsf{T}\mathtt{T}_i\mathbf{l}''. \tag{15.2}$$

Definition 15.1. The set of three matrices $\{\mathtt{T}_1, \mathtt{T}_2, \mathtt{T}_3\}$ constitute the *trifocal tensor* in matrix notation.

We introduce a further notation[1]. Denoting the ensemble of the three matrices $\mathtt{T}_i$ by $[\mathtt{T}_1, \mathtt{T}_2, \mathtt{T}_3]$, or more briefly $[\mathtt{T}_i]$, this last relation may be written as

$$\mathbf{l}^\mathsf{T} = \mathbf{l}'^\mathsf{T}[\mathtt{T}_1, \mathtt{T}_2, \mathtt{T}_3]\mathbf{l}'' \tag{15.3}$$

where $\mathbf{l}'^\mathsf{T}[\mathtt{T}_1, \mathtt{T}_2, \mathtt{T}_3]\mathbf{l}''$ is understood to represent the vector $(\mathbf{l}'^\mathsf{T}\mathtt{T}_1\mathbf{l}'', \mathbf{l}'^\mathsf{T}\mathtt{T}_2\mathbf{l}'', \mathbf{l}'^\mathsf{T}\mathtt{T}_3\mathbf{l}'')$.

Of course there is no intrinsic difference between the three views, and so by analogy with (15.3) there will exist similar relations $\mathbf{l}'^\mathsf{T} = \mathbf{l}^\mathsf{T}[\mathtt{T}'_i]\mathbf{l}''$ and $\mathbf{l}''^\mathsf{T} = \mathbf{l}^\mathsf{T}[\mathtt{T}''_i]\mathbf{l}'$. The three tensors $[\mathtt{T}_i]$, $[\mathtt{T}'_i]$ and $[\mathtt{T}''_i]$ exist, but are distinct. In fact, although all three tensors may be computed from any one of them, there is no very simple relationship between them. Thus, in fact there are three trifocal tensors existing for a given triple of views. Usually one will be content to consider only one of them. However, a method of computing the other trifocal tensors $[\mathtt{T}'_i]$ and $[\mathtt{T}''_i]$ given $[\mathtt{T}_i]$ is outlined in exercise (viii) on page 389.

Note that (15.3) is a relationship between image coordinates only, not involving 3D coordinates. Hence (as remarked previously), although it was derived under the assumption of a canonical camera set (that is $\mathtt{P} = [\mathtt{I} \mid \mathbf{0}]$), the value of the matrix elements $[\mathtt{T}_i]$ is independent of the form of the cameras. The particular simple formula (15.1) for the trifocal tensor given the camera matrices holds only in the case where

[1] This notation is somewhat cumbersome, and its meaning is not quite self-evident. It is for this reason that tensor notation is introduced in section 15.2.

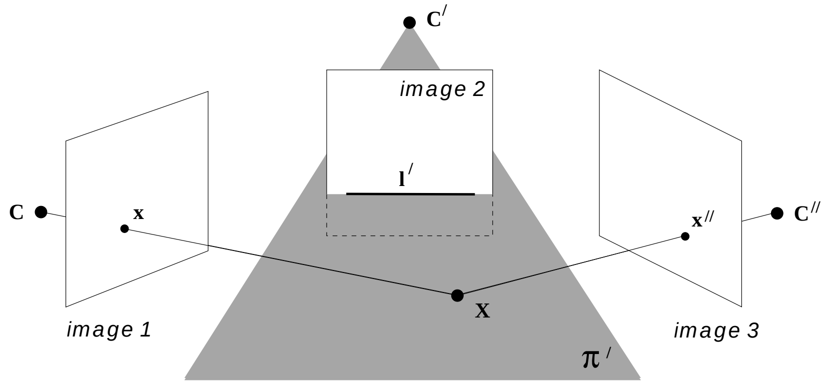


Fig. 15.2. **Point transfer.** *A line* $\mathbf{l}'$ *in the second view back-projects to a plane* $\boldsymbol{\pi}'$ *in 3-space. A point* $\mathbf{x}$ *in the first image defines a ray in 3-space which intersects* $\boldsymbol{\pi}'$ *in the point* $\mathbf{X}$. *This point* $\mathbf{X}$ *is then imaged as the point* $\mathbf{x}''$ *in the third view. Thus, any line* $\mathbf{l}'$ *induces a homography between the first and third views, defined by its back-projected plane* $\boldsymbol{\pi}'$.

$\mathtt{P} = [\mathtt{I} \mid \mathbf{0}]$, but a general formula (17.12–*p*415) for the trifocal tensor corresponding to any three cameras will be derived later.

Degrees of freedom. The trifocal tensor consists of three 3×3 matrices, and thus has 27 elements. There are therefore 26 independent ratios apart from the (common) overall scaling of the matrices. However, the tensor has only 18 independent degrees of freedom. In other words once 18 parameters are specified, all 27 elements of the tensor are determined up to a common scale. The number of degrees of freedom may be computed as follows. Each of 3 camera matrices has 11 degrees of freedom, which makes 33 in total. However, 15 degrees of freedom must be subtracted to account for the projective world frame, thus leaving 18 degrees of freedom. The tensor therefore satisfies $26 - 18 = 8$ independent algebraic constraints. We return to this point in chapter 16.

15.1.1 Homographies induced by a plane

A fundamental geometric property encoded in the trifocal tensor is the homography between the first view and the third induced by a line in the second image. This is illustrated in figure 15.2 and figure 15.3. A line in the second view defines (by back-projection) a plane in 3-space, and this plane induces a homography between the first and third views.

We now derive the algebraic representation of this geometry in terms of the trifocal tensor. The homography map between the first and third images, defined by the plane $\boldsymbol{\pi}'$ in figure 15.2 and figure 15.3, may be written as $\mathbf{x}'' = \mathtt{H}\mathbf{x}$ and (2.6–*p*36) $\mathbf{l} = \mathtt{H}^{\mathsf{T}}\mathbf{l}''$ respectively. Notice that the three lines $\mathbf{l}$, $\mathbf{l}'$ and $\mathbf{l}''$ in figure 15.3 are a corresponding line triple, the projections of the 3D line $\mathbf{L}$. Therefore, they satisfy the line incidence relationship $l_i = \mathbf{l}'^{\mathsf{T}}\mathtt{T}_i\mathbf{l}''$ of (15.2). Comparison of this formula and $\mathbf{l} = \mathtt{H}^{\mathsf{T}}\mathbf{l}''$ shows that

$$\mathtt{H} = [\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3] \text{ with } \mathbf{h}_i = \mathtt{T}_i^{\mathsf{T}}\mathbf{l}'.$$

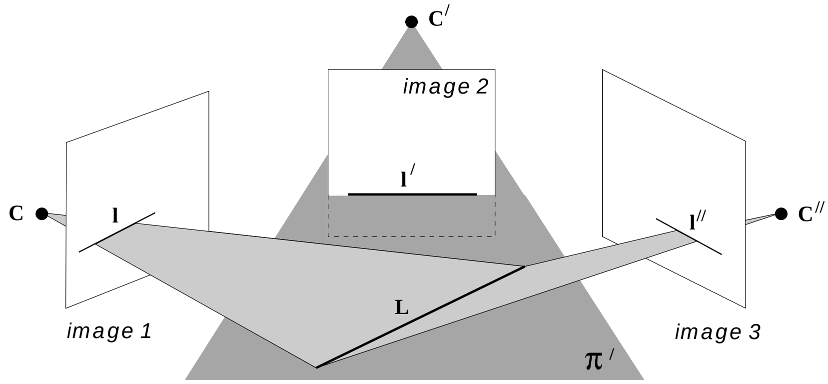


Fig. 15.3. **Line transfer.** *The action on lines of the homography defined by figure 15.2 may similarly be visualized geometrically. A line,* $\mathbf{l}$*, in the first image defines a plane in 3-space, which intersects* $\boldsymbol{\pi}'$ *in the line* $\mathbf{L}$*. This line* $\mathbf{L}$ *is then imaged as the line* $\mathbf{l}''$ *in the third view.*

Thus, $\mathtt{H}$ defined by the above formula represents the (point) homography $\mathtt{H}_{13}$ between views one and three specified by the line $\mathbf{l}'$ in view two.

The second and third views play similar roles, and the homography between the first and second views defined by a line in the third can be derived in a similar manner. These ideas are formalized in the following result.

Result 15.2. *The homography from the first to the third image induced by a line* $\mathbf{l}'$ *in the second image (see figure 15.2) is given by* $\mathbf{x}'' = \mathtt{H}_{13}(\mathbf{l}')\,\mathbf{x}$*, where*

$$\mathtt{H}_{13}(\mathbf{l}') = [\mathtt{T}_1^{\mathsf{T}}, \mathtt{T}_2^{\mathsf{T}}, \mathtt{T}_3^{\mathsf{T}}]\mathbf{l}'.$$

Similarly, a line $\mathbf{l}''$ *in the third image defines a homography* $\mathbf{x}' = \mathtt{H}_{12}(\mathbf{l}'')\,\mathbf{x}$ *from the first to the second views, given by*

$$\mathtt{H}_{12}(\mathbf{l}'') = [\mathtt{T}_1, \mathtt{T}_2, \mathtt{T}_3]\mathbf{l}''.$$

Once this mapping is understood the algebraic properties of the tensor are straightforward and can easily be generated. In the following section we deduce a number of incidence relations between points and lines based on (15.3) and result 15.2.

15.1.2 Point and line incidence relations

It is easy to deduce various linear relationships between lines and points in three images involving the trifocal tensor. We have seen one such relationship already, namely (15.3). This relation holds only up to scale since it involves homogeneous quantities. We may eliminate the scale factor by taking the vector cross product of both sides, which must be zero. This leads to the formula

$$(\mathbf{l}'^{\mathsf{T}}[\mathtt{T}_1, \mathtt{T}_2, \mathtt{T}_3]\mathbf{l}'')[\mathbf{l}]_\times = \mathbf{0}^{\mathsf{T}}, \tag{15.4}$$

where we have used the matrix $[\mathbf{l}]_\times$ to denote the cross product (see (A4.5–*p*581)), or more briefly $(\mathbf{l}'^{\mathsf{T}}[\mathtt{T}_i]\mathbf{l}'')[\mathbf{l}]_\times = \mathbf{0}^{\mathsf{T}}$. Note the symmetry between $\mathbf{l}'$ and $\mathbf{l}''$ – swapping the roles of these two lines is accounted for by transposing each $\mathtt{T}_i$, resulting in a relation $(\mathbf{l}''^{\mathsf{T}}[\mathtt{T}_i^{\mathsf{T}}]\mathbf{l}')[\mathbf{l}]_\times = \mathbf{0}^{\mathsf{T}}$.

Consider again figure 15.3. Now, a point $\mathbf{x}$ on the line $\mathbf{l}$ must satisfy $\mathbf{x}^{\mathsf{T}}\mathbf{l} = \sum_i x^i l_i = 0$ (using upper indices for the point coordinates, foreshadowing the use of tensor notation). Since $l_i = \mathbf{l}'^{\mathsf{T}}\mathtt{T}_i\mathbf{l}''$, this may be written as

$$\mathbf{l}'^{\mathsf{T}}(\sum_i x^i \mathtt{T}_i)\mathbf{l}'' = 0 \tag{15.5}$$

(note that $(\sum_i x^i\mathtt{T}_i)$ is simply a 3×3 matrix). This is an incidence relation in the first image: the relationship will hold for a point–line–line correspondence – that is whenever some 3D line $\mathbf{L}$ maps to $\mathbf{l}'$ and $\mathbf{l}''$ in the second and third images, and to a line passing through $\mathbf{x}$ in the first image. An important equivalent definition of a point–line–line correspondence for which (15.5) holds results from *an incidence relation in 3-space* – there exists a 3D point $\mathbf{X}$ mapping to $\mathbf{x}$ in the first image, and to points on the lines $\mathbf{l}'$ and $\mathbf{l}''$ in the second and third images as shown in figure 15.4(a).

From result 15.2 we may obtain relations involving points $\mathbf{x}'$ and $\mathbf{x}''$ in the second and third images. Consider a point–line–point correspondence as in figure 15.4(b) so that

$$\mathbf{x}'' = \mathtt{H}_{13}(\mathbf{l}')\,\mathbf{x} = [\mathtt{T}_1^{\mathsf{T}}\mathbf{l}', \mathtt{T}_2^{\mathsf{T}}\mathbf{l}', \mathtt{T}_3^{\mathsf{T}}\mathbf{l}']\,\mathbf{x} = (\sum_i x^i\mathtt{T}_i^{\mathsf{T}})\mathbf{l}'$$

which is valid for any line $\mathbf{l}'$ passing through $\mathbf{x}'$ in the second image. The homogeneous scale factor may be eliminated by (post-)multiplying the transpose of both sides by $[\mathbf{x}'']_\times$ to give

$$\mathbf{x}''^{\mathsf{T}}[\mathbf{x}'']_\times = \mathbf{l}'^{\mathsf{T}}(\sum_i x^i\mathtt{T}_i)[\mathbf{x}'']_\times = \mathbf{0}^{\mathsf{T}}, \tag{15.6}$$

A similar analysis may be undertaken with the roles of the second and third images swapped.

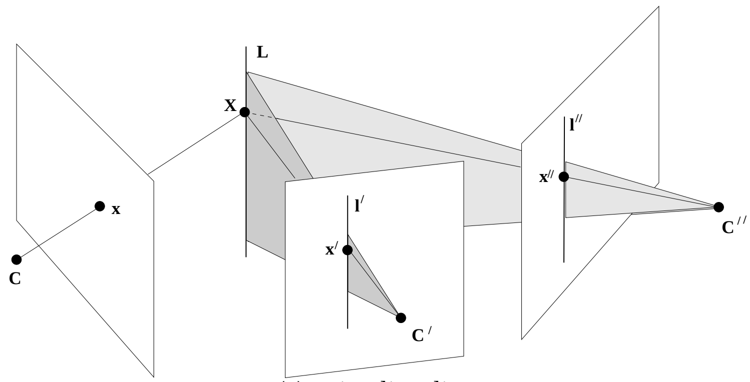
Finally, for a 3-point correspondence as shown in figure 15.4(c), there is a relation

$$[\mathbf{x}']_\times(\sum_i x^i\mathtt{T}_i)[\mathbf{x}'']_\times = \mathtt{0}_{3\times 3}. \tag{15.7}$$

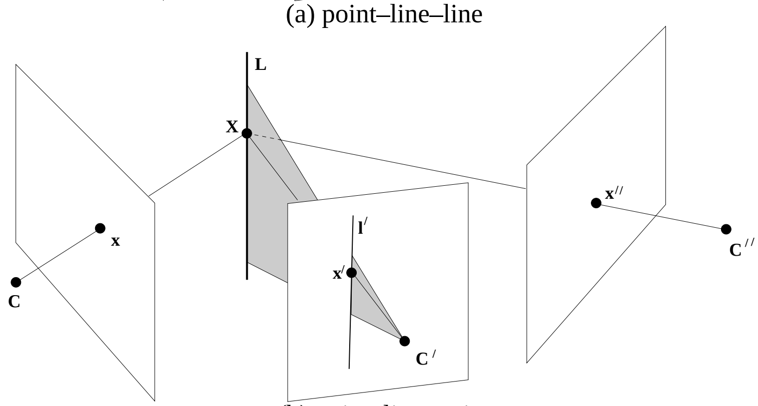
Proof. The line $\mathbf{l}'$ in (15.6) passes through $\mathbf{x}'$ and so may be written as $\mathbf{l}' = \mathbf{x}' \times \mathbf{y}' = [\mathbf{x}']_\times\mathbf{y}'$ for some point $\mathbf{y}'$ on $\mathbf{l}'$. Consequently, from (15.6) $\mathbf{l}'^{\mathsf{T}}(\sum_i x^i\mathtt{T}_i)[\mathbf{x}'']_\times = \mathbf{y}'^{\mathsf{T}}[\mathbf{x}']_\times(\sum_i x^i\mathtt{T}_i)[\mathbf{x}'']_\times = \mathbf{0}^{\mathsf{T}}$. However, the relation (15.6) is true for *all* lines $\mathbf{l}'$ through $\mathbf{x}'$ and so is independent of $\mathbf{y}'$. The relation (15.7) then follows. □

The various relationships between lines and points in three views are summarized in table 15.1, and their properties are investigated further in section 15.2.1, once tensor notation has been introduced. Note that there are no relations listed for point–line–line correspondence in which the point is in the second or third view. Such simple relations do not exist in terms of the trifocal tensor in which the first view is the special view.

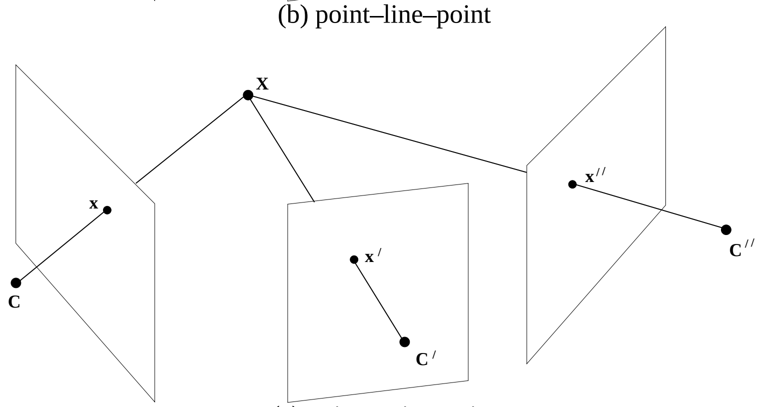
It is also worth noting that satisfying an image incidence relation does not guarantee incidence in 3-space, as illustrated in figure 15.5.



(a) point–line–line



(b) point–line–point



(c) point–point–point

Fig. 15.4. **Incidence relations.** *(a) Consider a 3-view point correspondence* $\mathbf{x} \leftrightarrow \mathbf{x}' \leftrightarrow \mathbf{x}''$. *If* $\mathbf{l}'$ *and* $\mathbf{l}''$ *are* **any** *two lines through* $\mathbf{x}'$ *and* $\mathbf{x}''$ *respectively, then* $\mathbf{x} \leftrightarrow \mathbf{l}' \leftrightarrow \mathbf{l}''$ *forms a point–line–line correspondence, corresponding to a 3D line* $\mathbf{L}$. *Consequently, (15.5) holds for any choice of lines* $\mathbf{l}'$ *through* $\mathbf{x}'$ *and* $\mathbf{l}''$ *through* $\mathbf{x}''$. *(b) The space point* $\mathbf{X}$ *is incident with the space line* $\mathbf{L}$. *This defines an incidence relation* $\mathbf{x} \leftrightarrow \mathbf{l}' \leftrightarrow \mathbf{x}''$ *between their images. (c) The correspondence* $\mathbf{x} \leftrightarrow \mathbf{x}' \leftrightarrow \mathbf{x}''$ *arising from the image of a space point* $\mathbf{X}$.

We now begin to extract the two-view geometry, the epipoles and fundamental matrix, from the trifocal tensor.

15.1.3 Epipolar lines

A special case of a point–line–line correspondence occurs when the plane $\boldsymbol{\pi}'$ back-projected from $\mathbf{l}'$ is an epipolar plane with respect to the first two cameras, and hence

(i) Line–line–line correspondence

$$\mathbf{l}'^{\mathsf{T}}[\mathtt{T}_1, \mathtt{T}_2, \mathtt{T}_3]\mathbf{l}'' = \mathbf{l}^{\mathsf{T}} \quad \text{or} \quad \left(\mathbf{l}'^{\mathsf{T}}[\mathtt{T}_1, \mathtt{T}_2, \mathtt{T}_3]\mathbf{l}''\right)[\mathbf{l}]_\times = \mathbf{0}^{\mathsf{T}}$$

(ii) Point–line–line correspondence

$$\mathbf{l}'^{\mathsf{T}}\Big(\sum_i x^i \mathtt{T}_i\Big)\mathbf{l}'' = 0 \ \text{ for a correspondence } \mathbf{x} \leftrightarrow \mathbf{l}' \leftrightarrow \mathbf{l}''$$

(iii) Point–line–point correspondence

$$\mathbf{l}'^{\mathsf{T}}\Big(\sum_i x^i \mathtt{T}_i\Big)[\mathbf{x}'']_\times = \mathbf{0}^{\mathsf{T}} \ \text{ for a correspondence } \mathbf{x} \leftrightarrow \mathbf{l}' \leftrightarrow \mathbf{x}''$$

(iv) Point–point–line correspondence

$$[\mathbf{x}']_\times\Big(\sum_i x^i \mathtt{T}_i\Big)\mathbf{l}'' = \mathbf{0} \ \text{ for a correspondence } \mathbf{x} \leftrightarrow \mathbf{x}' \leftrightarrow \mathbf{l}''$$

(v) Point–point–point correspondence

$$[\mathbf{x}']_\times\Big(\sum_i x^i \mathtt{T}_i\Big)[\mathbf{x}'']_\times = 0_{3\times 3}$$

Table 15.1. *Summary of trifocal tensor incidence relations using matrix notation.*

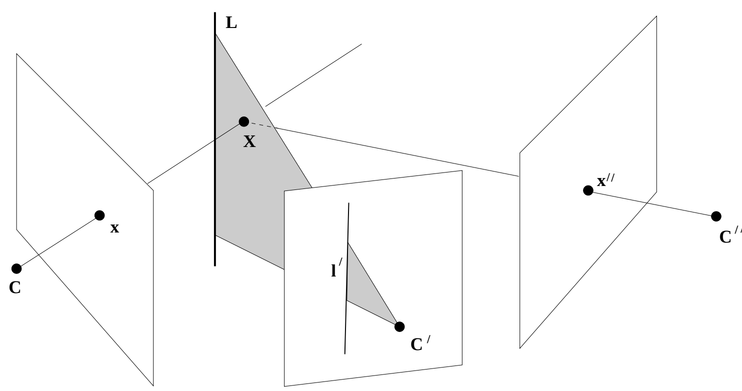


Fig. 15.5. **Non-incident configuration.** *The imaged points and lines of this configuration satisfy the point–line–point incidence relation of table 15.1. However, the space point* $\mathbf{X}$ *and line* $\mathbf{L}$ *are not incident. Compare with figure 15.4.*

passes through the camera centre $\mathbf{C}$ of the first camera. Suppose $\mathbf{X}$ is a point on the plane $\boldsymbol{\pi}'$; then the ray defined by $\mathbf{X}$ and $\mathbf{C}$ lies in this plane, and $\mathbf{l}'$ is the epipolar line corresponding to the point $\mathbf{x}$, the image of $\mathbf{X}$. This is shown in figure 15.6.

The plane $\boldsymbol{\pi}''$ back-projected from a line $\mathbf{l}''$ in the third image will intersect the plane $\boldsymbol{\pi}'$ in a line $\mathbf{L}$. Further, since the ray corresponding to $\mathbf{x}$ lies entirely in the plane $\boldsymbol{\pi}'$ it must intersect the line $\mathbf{L}$. This gives a 3-way intersection between the ray and planes back-projected from point $\mathbf{x}$ and lines $\mathbf{l}'$ and $\mathbf{l}''$, and so they constitute a point–line–line correspondence, satisfying $\mathbf{l}'^{\mathsf{T}}(\sum_i x^i \mathtt{T}_i)\mathbf{l}'' = 0$. The important point now is that this is true for *any* line $\mathbf{l}''$, and it follows that $\mathbf{l}'^{\mathsf{T}}(\sum_i x^i \mathtt{T}_i) = \mathbf{0}^{\mathsf{T}}$. The same argument holds with the roles of $\mathbf{l}'$ and $\mathbf{l}''$ reversed. To summarize:

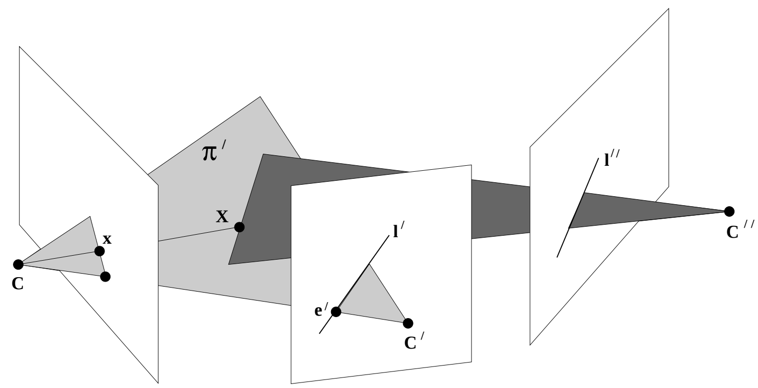


Fig. 15.6. *If the plane π' defined by $\mathbf{l}'$ is an epipolar plane for the first two views, then any line $\mathbf{l}''$ in the third view gives a point–line–line incidence.*

Result 15.3. *If $\mathbf{x}$ is a point and $\mathbf{l}'$ and $\mathbf{l}''$ are the corresponding epipolar lines in the second and third images, then*

$$\mathbf{l}'^{\mathsf{T}}(\sum_i x^i \mathtt{T}_i) = \mathbf{0}^{\mathsf{T}} \;\; \textit{and} \;\; (\sum_i x^i \mathtt{T}_i)\mathbf{l}'' = \mathbf{0}.$$

Consequently, the epipolar lines $\mathbf{l}'$ and $\mathbf{l}''$ corresponding to $\mathbf{x}$ may be computed as the left and right null-vectors of the matrix $\sum_i x^i \mathtt{T}_i$.

As the point $\mathbf{x}$ varies, the corresponding epipolar lines vary, but all epipolar lines in one image pass through the epipole. Thus, one may compute this epipole by computing the intersection of the epipolar lines for varying values of $\mathbf{x}$. Three convenient choices of $\mathbf{x}$ are the points represented by homogeneous coordinates $(1,0,0)^{\mathsf{T}}$, $(0,1,0)^{\mathsf{T}}$ and $(0,0,1)^{\mathsf{T}}$, with $\sum_i x^i \mathtt{T}_i$ equal to $\mathtt{T}_1$, $\mathtt{T}_2$ and $\mathtt{T}_3$ respectively for these three choices of $\mathbf{x}$. From this we deduce the following important result:

Result 15.4. *The epipole $\mathbf{e}'$ in the second image is the common intersection of the epipolar lines represented by the left null-vectors of the matrices $\mathtt{T}_i, i = 1, \ldots, 3$. Similarly the epipole $\mathbf{e}''$ is the common intersection of lines represented by the right null-vectors of the $\mathtt{T}_i$.*

Note that the epipoles involved here are the epipoles in the second and third images corresponding to the first image centre $\mathbf{C}$.

The usefulness of this result may not be apparent at present. However, it will be seen below that it is an important step in computing the camera matrices from the trifocal tensor, and in chapter 16 in the accurate computation of the trifocal tensor.

Algebraic properties of the $\mathtt{T}_i$ matrices. This section has established a number of algebraic properties of the $\mathtt{T}_i$ matrices. We summarize these here:

- Each matrix $\mathtt{T}_i$ has rank 2. This is evident from (15.1) since $\mathtt{T}_i = \mathbf{a}_i\mathbf{e}''^{\mathsf{T}} - \mathbf{e}'\mathbf{b}_i^{\mathsf{T}}$ is the sum of two outer products.
- The right null-vector of $\mathtt{T}_i$ is $\mathbf{l}''_i = \mathbf{e}'' \times \mathbf{b}_i$, and is the epipolar line in the third view for the point $\mathbf{x} = (1,0,0)^{\mathsf{T}}$, $(0,1,0)^{\mathsf{T}}$ or $(0,0,1)^{\mathsf{T}}$, as $i = 1, 2$ or 3 respectively.

- The epipole $\mathbf{e}''$ is the common intersection of the epipolar lines $\mathbf{l}''_i$ for $i = 1, 2, 3$.
- The left null-vector of $\mathtt{T}_i$ is $\mathbf{l}'_i = \mathbf{e}' \times \mathbf{a}_i$, and is the epipolar line in the second view for the point $\mathbf{x} = (1, 0, 0)^\mathsf{T}$, $(0, 1, 0)^\mathsf{T}$ or $(0, 0, 1)^\mathsf{T}$, as $i = 1, 2$ or 3 respectively.
- The epipole $\mathbf{e}'$ is the common intersection of the epipolar lines $\mathbf{l}'_i$ for $i = 1, 2, 3$.
- The *sum* of the matrices $\mathtt{M}(\mathbf{x}) = (\sum_i x^i \mathtt{T}_i)$ also has rank 2. The right null-vector of $\mathtt{M}(\mathbf{x})$ is the epipolar line $\mathbf{l}''$ of $\mathbf{x}$ in the third view, and its left null-vector is the epipolar line $\mathbf{l}'$ of $\mathbf{x}$ in the second view.

It's worth emphasizing again that although a particular canonical form of the camera matrices $\mathtt{P}, \mathtt{P}'$ and $\mathtt{P}''$ is used in the derivation, the epipolar properties of the $\mathtt{T}_i$ matrices are independent of this choice.

15.1.4 Extracting the fundamental matrices

It is simple to compute the fundamental matrices $\mathtt{F}_{21}$ and $\mathtt{F}_{31}$ between the first[1] and the other views from the trifocal tensor. It was seen in section 9.2.1(*p*242) that the epipolar line corresponding to some point can be derived by transferring the point to the other view via a homography and joining the transferred point to the epipole. Consider a point $\mathbf{x}$ in the first view. According to figure 15.2 and result 15.2, a line $\mathbf{l}''$ in the third view induces a homography from the first to the second view given by $\mathbf{x}' = ([\mathtt{T}_1, \mathtt{T}_2, \mathtt{T}_3]\mathbf{l}'')\,\mathbf{x}$. The epipolar line corresponding to $\mathbf{x}$ is then found by joining $\mathbf{x}'$ to the epipole $\mathbf{e}'$. This gives $\mathbf{l}' = [\mathbf{e}']_\times ([\mathtt{T}_1, \mathtt{T}_2, \mathtt{T}_3]\mathbf{l}'')\,\mathbf{x}$, from which it follows that

$$\mathtt{F}_{21} = [\mathbf{e}']_\times [\mathtt{T}_1, \mathtt{T}_2, \mathtt{T}_3]\mathbf{l}''.$$

This formula holds for any vector $\mathbf{l}''$, but it is important to choose $\mathbf{l}''$ to avoid the degenerate condition where $\mathbf{l}''$ lies in the null-space of any of the $\mathtt{T}_i$. A good choice is $\mathbf{e}''$ since as has been seen $\mathbf{e}''$ is perpendicular to the right null-space of each $\mathtt{T}_i$. This gives the formula

$$\mathtt{F}_{21} = [\mathbf{e}']_\times [\mathtt{T}_1, \mathtt{T}_2, \mathtt{T}_3]\mathbf{e}''. \tag{15.8}$$

A similar formula holds for $\mathtt{F}_{31} = [\mathbf{e}'']_\times [\mathtt{T}_1^\mathsf{T}, \mathtt{T}_2^\mathsf{T}, \mathtt{T}_3^\mathsf{T}]\mathbf{e}'$.

15.1.5 Retrieving the camera matrices

It was remarked that the trifocal tensor, since it expresses a relationship between image entities only, is independent of 3D projective transformations. Conversely, this implies that the camera matrices may be computed from the trifocal tensor only up to a projective ambiguity. It will now be shown how this may be done.

Just as in the case of reconstruction from two views, because of the projective ambiguity, the first camera may be chosen as $\mathtt{P} = [\mathtt{I} \mid \mathbf{0}]$. Now, since $\mathtt{F}_{21}$ is known (from (15.8)), we can make use of result 9.9(*p*254) to derive the form of the second camera as

$$\mathtt{P}' = [[\mathtt{T}_1, \mathtt{T}_2, \mathtt{T}_3]\mathbf{e}'' \mid \mathbf{e}']$$

and the camera pair $\{\mathtt{P}, \mathtt{P}'\}$ then has the fundamental matrix $\mathtt{F}_{21}$. It might be thought

[1] The fundamental matrix $\mathtt{F}_{21}$ satisfies $\mathbf{x}'^\mathsf{T}\mathtt{F}_{21}\mathbf{x} = 0$ for corresponding points $\mathbf{x} \leftrightarrow \mathbf{x}'$. The subscript notation refers to figure 15.8.

Given the trifocal tensor written in matrix notation as $[\mathtt{T}_1, \mathtt{T}_2, \mathtt{T}_3]$.

(i) **Retrieve the epipoles** $\mathbf{e}', \mathbf{e}''$
Let $\mathbf{u}_i$ and $\mathbf{v}_i$ be the left and right null-vectors respectively of $\mathtt{T}_i$, i.e. $\mathbf{u}_i^\mathsf{T}\mathtt{T}_i = \mathbf{0}^\mathsf{T}$, $\mathtt{T}_i\mathbf{v}_i = \mathbf{0}$. Then the epipoles are obtained as the null-vectors to the following 3×3 matrices:

$$\mathbf{e}'^\mathsf{T}[\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3] = \mathbf{0} \text{ and } \mathbf{e}''^\mathsf{T}[\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3] = \mathbf{0}.$$

(ii) **Retrieve the fundamental matrices** $\mathtt{F}_{21}, \mathtt{F}_{31}$

$$\mathtt{F}_{21} = [\mathbf{e}']_\times[\mathtt{T}_1, \mathtt{T}_2, \mathtt{T}_3]\mathbf{e}'' \text{ and } \mathtt{F}_{31} = [\mathbf{e}'']_\times[\mathtt{T}_1^\mathsf{T}, \mathtt{T}_2^\mathsf{T}, \mathtt{T}_3^\mathsf{T}]\mathbf{e}'.$$

(iii) **Retrieve the camera matrices** $\mathtt{P}', \mathtt{P}''$ **(with** $\mathtt{P} = [\mathtt{I} \mid \mathbf{0}]$**)**
Normalize the epipoles to unit norm. Then

$$\mathtt{P}' = [[\mathtt{T}_1, \mathtt{T}_2, \mathtt{T}_3]\mathbf{e}'' \mid \mathbf{e}'] \text{ and } \mathtt{P}'' = [(\mathbf{e}''\mathbf{e}''^\mathsf{T} - \mathtt{I})[\mathtt{T}_1^\mathsf{T}, \mathtt{T}_2^\mathsf{T}, \mathtt{T}_3^\mathsf{T}]\mathbf{e}' \mid \mathbf{e}''].$$

Algorithm 15.1. **Summary of** $\mathtt{F}$ **and** $\mathtt{P}$ **retrieval from the trifocal tensor.** *Note,* $\mathtt{F}_{21}$ *and* $\mathtt{F}_{31}$ *are determined uniquely. However,* $\mathtt{P}'$ *and* $\mathtt{P}''$ *are determined only up to a common projective transformation of 3-space.*

that the third camera could be chosen in a similar manner as $\mathtt{P}'' = [[\mathtt{T}_1^\mathsf{T}, \mathtt{T}_2^\mathsf{T}, \mathtt{T}_3^\mathsf{T}]\mathbf{e}' \mid \mathbf{e}'']$, but *this is incorrect*. This is because the two camera pairs $\{\mathtt{P}, \mathtt{P}'\}$ and $\{\mathtt{P}, \mathtt{P}''\}$ do not necessarily define the same projective world frame; although each pair is correct by itself, the triple $\{\mathtt{P}, \mathtt{P}', \mathtt{P}''\}$ is inconsistent.

The third camera cannot be chosen independently of the projective frame of the first two. To see this, suppose the camera pair $\{\mathtt{P}, \mathtt{P}'\}$ is chosen and points $\mathbf{X}_i$ reconstructed from their image correspondences $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$. Then the coordinates of $\mathbf{X}_i$ are specified in the projective world frame defined by by $\{\mathtt{P}, \mathtt{P}'\}$, and a consistent camera $\mathtt{P}''$ may be computed from the correspondences $\mathbf{X}_i \leftrightarrow \mathbf{x}''_i$. Clearly, $\mathtt{P}''$ depends on the frame defined by $\{\mathtt{P}, \mathtt{P}'\}$. However, it is not necessary to explicitly reconstruct 3D structure, a consistent camera triplet can be recovered from the trifocal tensor directly.

The pair of camera matrices $\mathtt{P} = [\mathtt{I} \mid \mathbf{0}]$ and $\mathtt{P}' = [[\mathtt{T}_1, \mathtt{T}_2, \mathtt{T}_3]\mathbf{e}'' \mid \mathbf{e}']$ are not the only ones compatible with the given fundamental matrix $\mathtt{F}_{21}$. According to (9.10–*p*256), the most general form for $\mathtt{P}'$ is

$$\mathtt{P}' = [[\mathtt{T}_1, \mathtt{T}_2, \mathtt{T}_3]\mathbf{e}'' + \mathbf{e}'\mathbf{v}^\mathsf{T} | \lambda\mathbf{e}']$$

for some vector $\mathbf{v}$ and scalar λ. A similar choice holds for $\mathtt{P}''$. To find a triple of camera matrices compatible with the trifocal tensor, we need to find the correct values of $\mathtt{P}'$ and $\mathtt{P}''$ from these families so as to be compatible with the form (15.1) of the trifocal tensor.

Because of the projective ambiguity, we are free to choose $\mathtt{P}' = [[\mathtt{T}_1, \mathtt{T}_2, \mathtt{T}_3]\mathbf{e}''|\mathbf{e}']$, thus $\mathbf{a}_i = \mathtt{T}_i\mathbf{e}''$. This choice fixes the projective world frame so that $\mathtt{P}''$ is now defined uniquely (up to scale). Then substituting into (15.1) (observing that $\mathbf{a}_4 = \mathbf{e}'$ and $\mathbf{b}_4 = \mathbf{e}''$)

$$\mathtt{T}_i = \mathtt{T}_i\mathbf{e}''\mathbf{e}''^\mathsf{T} - \mathbf{e}'\mathbf{b}_i^\mathsf{T}$$

from which it follows that $\mathbf{e}'\mathbf{b}_i^\mathsf{T} = \mathtt{T}_i(\mathbf{e}''\mathbf{e}''^\mathsf{T} - \mathtt{I})$. Since the scale may be chosen such that $\|\mathbf{e}'\| = \mathbf{e}'^{\mathsf{T}}\mathbf{e}' = 1$, we may multiply on the left by $\mathbf{e}'^{\mathsf{T}}$ and transpose to get

$$\mathbf{b}_i = (\mathbf{e}''\mathbf{e}''^{\mathsf{T}} - \mathtt{I})\mathtt{T}_i^{\mathsf{T}}\mathbf{e}'$$

so $\mathtt{P}'' = [(\mathbf{e}''\mathbf{e}''^{\mathsf{T}} - \mathtt{I})[\mathtt{T}_1^{\mathsf{T}}, \mathtt{T}_2^{\mathsf{T}}, \mathtt{T}_3^{\mathsf{T}}]\mathbf{e}'|\mathbf{e}'']$. A summary of the steps involved in extracting the camera matrices from the trifocal tensor is given in algorithm 15.1.

We have seen that the trifocal tensor may be computed from the three camera matrices, and that conversely the three camera matrices may be computed, up to projective equivalence, from the trifocal tensor. Thus, the trifocal tensor completely captures the three cameras up to projective equivalence.

15.2 The trifocal tensor and tensor notation

The style of notation that has been used up to now for the trifocal tensor is derived from the standard matrix–vector notation. Since a matrix has two indices only, it is possible to distinguish between the two indices using the devices of matrix transposition and right or left multiplication, and in dealing with matrices and vectors, one can do without writing the indices explicitly. Because the trifocal tensor has three indices, instead of the two indices that a matrix has, it becomes increasingly cumbersome to persevere with this style of matrix notation, and we now turn to using standard tensor notation when dealing with the trifocal tensor. For those unfamiliar with tensor notation a gentle introduction is given in appendix 1(*p*562). This appendix should be read before proceeding with this chapter.

Image points and lines are represented by homogeneous *column* and *row* 3-vectors, respectively, i.e. $\mathbf{x} = (x^1, x^2, x^3)^{\mathsf{T}}$ and $\mathbf{l} = (l_1, l_2, l_3)$. The ij-th entry of a matrix $\mathtt{A}$ is denoted by a^i_j, index i being the contravariant (row) index and j being the covariant (column) index. We observe the convention that indices repeated in the contravariant and covariant positions imply summation over the range $(1, \ldots, 3)$ of the index. For example, the equation $\mathbf{x}' = \mathtt{A}\mathbf{x}$ is equivalent to $x'^i = \sum_j a^i_j x^j$, which may be written $x'^i = a^i_j x^j$.

We begin with the definition of the trifocal tensor given in (15.1). Using tensor notation, this becomes

$$\mathcal{T}_i^{jk} = a^j_i b^k_4 - a^j_4 b^k_i. \tag{15.9}$$

The positions of the indices in $\mathcal{T}_i^{jk}$ (two contravariant and one covariant) are dictated by the positions of the indices on the right side of the equation. Thus, the trifocal tensor is a mixed contravariant–covariant tensor. In tensor notation, the basic incidence relation (15.3) becomes

$$l_i = l'_j l''_k \mathcal{T}_i^{jk}. \tag{15.10}$$

Note that when multiplying tensors the order of the entries does not matter, in contrast with standard matrix notation. For instance the right side of the above expression is

$$l'_j l''_k \mathcal{T}_i^{jk} = \sum_{j,k} l'_j l''_k \mathcal{T}_i^{jk} = \sum_{j,k} l'_j \mathcal{T}_i^{jk} l''_k = l'_j \mathcal{T}_i^{jk} l''_k \ .$$

Definition. The trifocal tensor $\mathcal{T}$ is a valency 3 tensor $\mathcal{T}_i^{jk}$ with two contravariant and one covariant indices. It is represented by a homogeneous $3 \times 3 \times 3$ array (i.e. 27 elements). It has 18 degrees of freedom.

Computation from camera matrices. If the canonical 3×4 camera matrices are

$$\mathtt{P} = [\mathtt{I} \mid \mathbf{0}], \quad \mathtt{P}' = [a_j^i], \quad \mathtt{P}'' = [b_j^i]$$

then

$$\mathcal{T}_i^{jk} = a_i^j b_4^k - a_4^j b_i^k.$$

See (17.12–*p*415) for computation from three general camera matrices.

Line transfer from corresponding lines in the second and third views to the first.

$$l_i = l'_j l''_k \mathcal{T}_i^{jk}$$

Transfer by a homography.

(i) **Point transfer from first to third view via a plane in the second**
The contraction $l'_j \mathcal{T}_i^{jk}$ is a homography mapping between the first and third views induced by a plane defined by the back-projection of the line $\mathbf{l}'$ in the second view.

$$x''^k = h_i^k x^i \text{ where } h_i^k = l'_j \, \mathcal{T}_i^{jk}$$

(ii) **Point transfer from first to second view via a plane in the third**
The contraction $l''_k \mathcal{T}_i^{jk}$ is a homography mapping between the first and second views induced by a plane defined by the back-projection of the line $\mathbf{l}''$ in the third view.

$$x'^j = h_i^j x^i \text{ where } h_i^j = l''_k \, \mathcal{T}_i^{jk}$$

Table 15.2. *Definition and transfer properties of the trifocal tensor.*

The homography maps of figure 15.2 and figure 15.3 may be deduced from the incidence relation (15.10). In the case of the plane defined by back-projecting the line $\mathbf{l}'$,

$$l_i = l'_j l''_k \mathcal{T}_i^{jk} = l''_k (l'_j \mathcal{T}_i^{jk}) = l''_k \, h_i^k \text{ where } h_i^k = l'_j \mathcal{T}_i^{jk}$$

and h_i^k are the elements of the homography matrix $\mathtt{H}$. This homography maps points between the first and third view as

$$x''^k = h_i^k \, x^i.$$

Note that the homography is obtained from the tensor by contraction with a line (i.e. a summation over one contravariant (upper) index of the tensor, and the covariant (lower) index of the line), i.e. $\mathbf{l}'$ extracts a 3×3 matrix from the tensor – think of the trifocal tensor as an operator which takes a line and produces a homography matrix. Table 15.2 summarizes the definition and transfer properties of the trifocal tensor.

A pair of particularly important tensors are ϵ_{ijk} and its contravariant counterpart ϵ^{ijk}, defined in section A1.1(*p*563). This tensor is used to represent the vector product. For

(i) Line–line–line correspondence

$$(l_r \epsilon^{ris}) l'_j l''_k \mathcal{T}_i^{jk} = 0^s$$

(ii) Point–line–line correspondence

$$x^i l'_j l''_k \mathcal{T}_i^{jk} = 0$$

(iii) Point–line–point correspondence

$$x^i l'_j (x''^k \epsilon_{kqs}) \mathcal{T}_i^{jq} = 0_s$$

(iv) Point–point–line correspondence

$$x^i (x'^j \epsilon_{jpr}) l''_k \mathcal{T}_i^{pk} = 0_r$$

(v) Point–point–point correspondence

$$x^i (x'^j \epsilon_{jpr})(x''^k \epsilon_{kqs}) \mathcal{T}_i^{pq} = 0_{rs}$$

Table 15.3. *Summary of trifocal tensor incidence relations – the trilinearities.*

instance, the line joining two points x^i and y^j is equal to the cross product $x^i y^j \epsilon_{ijk} = l_k$, and the skew-symmetric matrix $[\mathbf{x}]_\times$ is written as $x^i \epsilon_{irs}$ in tensor notation. It is now relatively straightforward to write down the basic incidence results involving the trifocal tensor given in table 15.1. The results are summarized in table 15.3. In this table, a notation such as 0_r represents an array of zeros.

The form of the relations in table 15.3 is more easily understood if one observes that three indices i, j and k in $\mathcal{T}_i^{jk}$ correspond to entities in the first, second and third views respectively. Thus for instance a partial expression such as $l''_j \mathcal{T}_i^{jk}$ cannot occur, because the index j belongs to the second view, and hence does not belong on the line $\mathbf{l}''$ in the third view. Repeated indices (indicating summation) must occur once as a contravariant (upper) index and once as a covariant (lower) index. Thus, we cannot write $x'^j \mathcal{T}_i^{jk}$, since the index j occurs twice in the upper position. Think of the ϵ tensor as being used to raise or lower indices, for instance by replacing $\mathbf{l}'_j$ by $\mathbf{x}^i \epsilon_{ijk}$. However, this may not be done arbitrarily, as pointed out in exercise (x) on page 389.

15.2.1 The trilinearities

The incidence relations in table 15.3 are *trilinear* relations or *trilinearities* in the coordinates of the image elements (points and lines). *Tri-* since every monomial in the relation involves a coordinate from each of the three image elements involved; and *linear* because the relations are linear in each of the algebraic entities (i.e. the three "arguments" of the tensor). For example in the point–point–point relation, $x^i (x'^j \epsilon_{jpr})(x''^k \epsilon_{kqs}) \mathcal{T}_i^{pq} = 0_{rs}$, suppose both $\mathbf{x}_1$ and $\mathbf{x}_2$ satisfy the relation, then so does $\mathbf{x} = \alpha \mathbf{x}_1 + \beta \mathbf{x}_2$, i.e. the relation is linear in its first argument. Similarly, the relation is linear in the second and third argument. This multi-linearity is a standard property of tensors, and follows directly from the form $x^i l'_j l''_k \mathcal{T}_i^{jk} = 0$ which is a contraction of the tensor over all three of its indices (arguments).

We will now describe the point–point–point trilinearities in more detail. There are nine of these trilinearities arising from the three choices of r and s. Geometrically these trilinearities arise from special choices of the lines in the second and third image for the point–line–line relation (see figure 15.4(a)). Choosing $r = 1, 2$ or 3 corresponds to a line parallel to the image x-axis, parallel to the image y-axis, or through the image coordinate origin (the point $(0, 0, 1)^{\mathsf{T}}$), respectively. For example, choosing $r = 1$ and expanding $x'^j \epsilon_{jpr}$ results in

$$l'_p = x'^j \epsilon_{jp1} = (0, -x'^3, x'^2)$$

which is a horizontal line in the second view through $\mathbf{x}'$ (since points of the form $\mathbf{y}' = (x'^1 + \lambda, x'^2, x'^3)^{\mathsf{T}}$ satisfy $\mathbf{y}'^{\mathsf{T}}\mathbf{l}' = 0$ for any λ). Similarly, choosing $s = 2$ in the third view results in the vertical line through $\mathbf{x}''$

$$l''_q = x''^k \epsilon_{kq2} = (x''^3, 0, -x''^1)$$

and the trilinear point relation expands to

$$\begin{aligned} 0 &= x^i x'^j x''^k \, \epsilon_{jp1} \epsilon_{kq2} \mathcal{T}_i^{pq} \\ &= x^i [-x'^3 (x''^3 \, \mathcal{T}_i^{21} - x''^1 \, \mathcal{T}_i^{23}) + x'^2 (x''^3 \, \mathcal{T}_i^{31} - x''^1 \, \mathcal{T}_i^{33})]. \end{aligned}$$

Of these nine trilinearities, four are linearly independent. This means that from a basis of four trilinearities all nine can be generated by linear combinations. The four degrees of freedom may be traced back to those of the point-line-line relation $x^i l'_j l''_k \mathcal{T}_i^{jk} = 0$ and are counted as follows. There is a one-parameter family of lines through $\mathbf{x}''$ in the third view. If $\mathbf{m}''$ and $\mathbf{n}''$ are two members of this family, then any other line through $\mathbf{x}''$ can be obtained from a linear combination of these:

$$\mathbf{l}'' = \alpha \mathbf{m}'' + \beta \mathbf{n}''.$$

The incidence relation is linear in $\mathbf{l}''$, so that given

$$\begin{aligned} l'_j m''_k \mathcal{T}_i^{jk} x^i &= 0 \\ l'_j n''_k \mathcal{T}_i^{jk} x^i &= 0 \end{aligned}$$

then the incidence relation for any other line $\mathbf{l}''$ can be generated by a linear combination of these two. Consequently, there are only two linearly independent incidence relations for $\mathbf{l}''$. Similarly there is a one-parameter family of lines through $\mathbf{x}'$, and the incidence relation is also linear in lines $\mathbf{l}'$ through $\mathbf{x}'$. Thus, there are a total of four linearly independent incidence relations between a point in the first view and lines in the second and third.

The main virtue of the trilinearities is that they are linear, otherwise their properties are often subsumed by transfer, as described in the following section.

15.3 Transfer

Given three views of a scene and a pair of matched points in two views one may wish to determine the position of the point in the third view. Given sufficient information about the placement of the cameras, it is usually possible to determine the location of

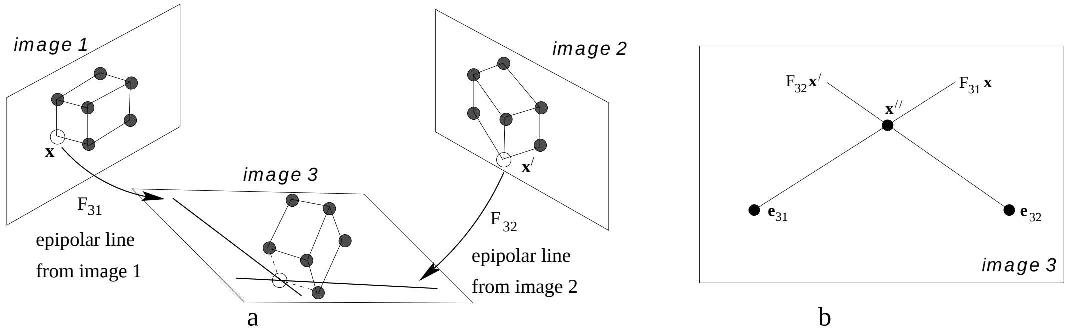


Fig. 15.7. **Epipolar transfer.** *(a) The image of* $\mathbf{X}$ *in the first two views is the correspondence* $\mathbf{x} \leftrightarrow \mathbf{x}'$. *The image of* $\mathbf{X}$ *in the third view may be computed by intersecting the epipolar lines* $\mathtt{F}_{31}\mathbf{x}$ *and* $\mathtt{F}_{32}\mathbf{x}'$. *(b) The configuration of the epipoles and transferred point* $\mathbf{x}''$ *as seen in the third image. Point* $\mathbf{x}''$ *is computed as the intersection of epipolar lines passing through the two epipoles* $\mathbf{e}_{31}$ *and* $\mathbf{e}_{32}$. *However, if* $\mathbf{x}''$ *lies on the line through the two epipoles, then its position cannot be determined. Points close to the line through the epipoles will be estimated with poor precision.*

the point in the third view without reference to image content. This is the point transfer problem. A similar transfer problem arises for lines.

In principle the problem can generally be solved given the cameras for the three views. Rays back-projected from corresponding points in the first and second view intersect and thus determine the 3D point. The position of the corresponding point in the third view is computed by projecting this 3D point onto the image. Similarly lines back-projected from the first and second image intersect in the 3D line, and the projection of this line in 3-space to the third image determines its image position.

15.3.1 Point transfer using fundamental matrices

The transfer problem may be solved using knowledge of the fundamental matrices only. Thus, suppose we know the three fundamental matrices $\mathtt{F}_{21}$, $\mathtt{F}_{31}$ and $\mathtt{F}_{32}$ relating the three views, and let points $\mathbf{x}$ and $\mathbf{x}'$ in the first two views be a matched pair. We wish to find the corresponding point $\mathbf{x}''$ in the third image.

The required point $\mathbf{x}''$ matches point $\mathbf{x}$ in the first image, and consequently must lie on the epipolar line corresponding to $\mathbf{x}$. Since we know $\mathtt{F}_{31}$, this epipolar line may be computed, and is equal to $\mathtt{F}_{31}\mathbf{x}$. By a similar argument, $\mathbf{x}''$ must lie on the epipolar line $\mathtt{F}_{32}\mathbf{x}'$. Taking the intersection of the epipolar lines gives

$$\mathbf{x}'' = (\mathtt{F}_{31}\mathbf{x}) \times (\mathtt{F}_{32}\mathbf{x}') \ .$$

See figure 15.7a.

Note that the fundamental matrix $\mathtt{F}_{21}$ is not used in this expression. The question naturally arises whether we can gain anything by knowledge of $\mathtt{F}_{21}$, and the answer is yes. In the presence of noise, the points $\mathbf{x} \leftrightarrow \mathbf{x}'$ will not form an exact matched pair, meaning that they will not satisfy the equation $\mathbf{x}'^{\mathsf{T}}\mathtt{F}_{21}\mathbf{x} = 0$ exactly. Given $\mathtt{F}_{21}$ one may use optimal triangulation as in algorithm 12.1(*p*318) to correct $\mathbf{x}$ and $\mathbf{x}'$, resulting in a pair $\hat{\mathbf{x}} \leftrightarrow \hat{\mathbf{x}}'$ that satisfies this relation. The transferred point may then be computed as $\mathbf{x}'' = (\mathtt{F}_{31}\hat{\mathbf{x}}) \times (\mathtt{F}_{32}\hat{\mathbf{x}}')$. This method of point transfer using the fundamental matrices will be called *epipolar transfer*.

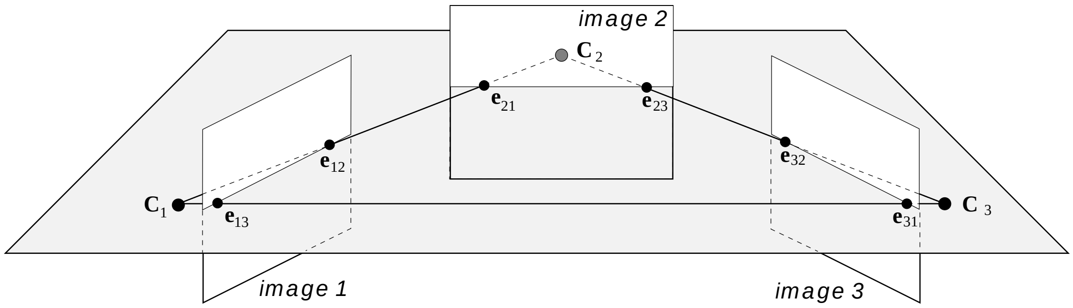


Fig. 15.8. **The trifocal plane** *is defined by the three camera centres. The notation for the epipoles is* $\mathbf{e}_{ij} = \mathtt{P}_i\mathbf{C}_j$. *Epipolar transfer fails for any point* $\mathbf{X}$ *on the trifocal plane. If the three camera centres are collinear then there is a one-parameter family of planes containing the three centres.*

Though at one time used for point transfer, epipolar transfer has a serious deficiency that rules it out as a practical method. This deficiency is due to the degeneracy that can be seen from figure 15.7(b): epipolar transfer fails when the two epipolar lines in the third image are coincident (and becomes increasingly ill-conditioned as the lines become less "transverse"). The degeneracy condition that $\mathbf{x}''$, $\mathbf{e}_{31}$ and $\mathbf{e}_{32}$ are collinear in the third image means that the camera centres $\mathbf{C}$ and $\mathbf{C}'$ and the 3D point $\mathbf{X}$ lie in a plane through the centre $\mathbf{C}''$ of the third camera; thus $\mathbf{X}$ lies on the *trifocal* plane defined by the three camera centres, see figure 15.8. Epipolar transfer will fail for points $\mathbf{X}$ lying on the trifocal plane and will be inaccurate for points lying near that plane. Note, in the special case that the three camera centres are collinear the trifocal plane is not uniquely defined, and epipolar transfer fails for all points. In this case $\mathbf{e}_{31} = \mathbf{e}_{32}$.

15.3.2 Point transfer using the trifocal tensor

The degeneracy of epipolar transfer is avoided by use of the trifocal tensor. Consider a correspondence $\mathbf{x} \leftrightarrow \mathbf{x}'$. If a line $\mathbf{l}'$ passing through the point $\mathbf{x}'$ is chosen in the second view, then the corresponding point $\mathbf{x}''$ may be computed by transferring the point $\mathbf{x}$ from the first to the third view using $x''^k = x^i l'_j \mathcal{T}_i^{jk}$, from table 15.2. It is clear from figure 15.4(*p*371)(b) that this transfer is not degenerate for general points $\mathbf{X}$ lying on the trifocal plane.

However, note from result 15.3 and figure 15.6 that if $\mathbf{l}'$ is the epipolar line corresponding to $\mathbf{x}$, then $x^i l'_j \mathcal{T}_i^{jk} = 0^k$, so the point $\mathbf{x}''$ is undefined. Consequently, the choice of line $\mathbf{l}'$ is important. To avoid choosing only an epipolar line, one possibility is to use two or three different lines passing through $\mathbf{x}'$, namely $l'_{jp} = x'^r \epsilon_{rjp}$ for the three choices of $p = 1, \ldots, 3$. For each such line, one computes the value of $\mathbf{x}''$ and retains the one that has the largest norm (i.e. is furthest from being zero). An alternative method entirely for finding $\mathbf{x}''$ is as the least-squares solution of the system of linear equations $x^i(x'^j \epsilon_{jpr})(x''^k \epsilon_{kqs})\mathcal{T}_i^{pq} = 0_{rs}$, but this method is probably an overkill.

The method we recommend is the following. Before attempting to compute the point $\mathbf{x}''$ transferred from a pair of points $\mathbf{x} \leftrightarrow \mathbf{x}'$, first correct the pair of points using the fundamental matrix $\mathtt{F}_{21}$, as described above in the case of epipolar transfer. If $\hat{\mathbf{x}}$ and $\hat{\mathbf{x}}'$

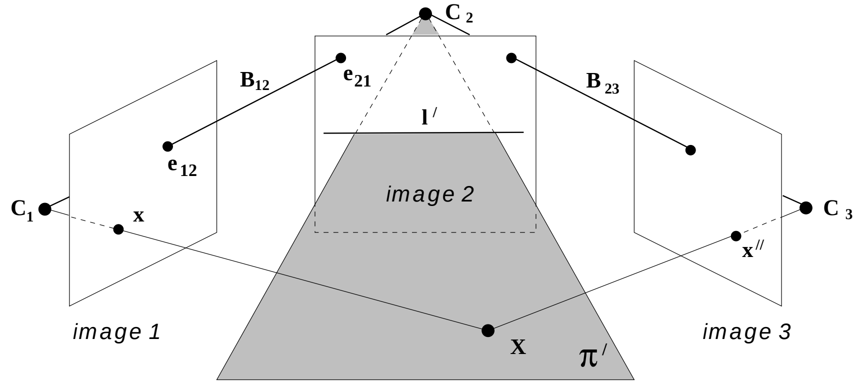


Fig. 15.9. **Degeneracy for point transfer using the trifocal tensor.** *The 3D point* $\mathbf{X}$ *is defined by the intersection of the ray through* $\mathbf{x}$ *with the plane* $\boldsymbol{\pi}'$. *A point* $\mathbf{X}$ *on the baseline* B_{12} *between the first and second views cannot be defined in this manner. So a 3D point on the line* B_{12} *cannot be transferred to the third view via a homography defined by a line in the second view. Note that a point on the line* B_{12} *projects to* $\mathbf{e}_{12}$ *in the first image and* $\mathbf{e}_{21}$ *in the second image. Apart from the line* B_{12} *any point can be transferred. In particular there is not a degeneracy problem for points on the baseline* B_{23}, *between views two and three, or for any other point on the trifocal plane.*

are an exact match, then the transferred point $x''^k = \hat{x}^i l'_j \mathcal{T}_i^{jk}$ does not depend on the line $\mathbf{l}'$ chosen passing through $\hat{\mathbf{x}}'$ (as long as it is not the epipolar line). This may be verified geometrically by referring to figure 15.2(*p*368). A good choice is always given by the line perpendicular to $\mathtt{F}_{21}\hat{\mathbf{x}}$.

To summarize, a measured correspondence $\mathbf{x} \leftrightarrow \mathbf{x}'$ is transferred by the following steps:

(i) Compute $\mathtt{F}_{21}$ from the trifocal tensor (by the method given in algorithm 15.1), and correct $\mathbf{x} \leftrightarrow \mathbf{x}'$ to the exact correspondence $\hat{\mathbf{x}} \leftrightarrow \hat{\mathbf{x}}'$ using algorithm 12.1-(*p*318).
(ii) Compute the line $\mathbf{l}'$ through $\hat{\mathbf{x}}'$ and perpendicular to $\mathbf{l}'_e = \mathtt{F}_{21}\hat{\mathbf{x}}$. If $\mathbf{l}'_e = (l_1, l_2, l_3)^\mathsf{T}$ and $\hat{\mathbf{x}}' = (\hat{x}_1, \hat{x}_2, 1)^\mathsf{T}$, then $\mathbf{l}' = (l_2, -l_1, -\hat{x}_1 l_2 + \hat{x}_2 l_1)^\mathsf{T}$.
(iii) The transferred point is $x''^k = \hat{x}^i l'_j \mathcal{T}_i^{jk}$.

Degenerate configurations. Consider transfer to the third view via a plane, as shown in figure 15.9. The 3D point $\mathbf{X}$ is only undefined if it lies on the baseline joining the first and second camera centres. This is because rays through $\mathbf{x}$ and $\mathbf{x}'$ are collinear for such 3D points and so their intersection is not defined. In such a case, the points $\mathbf{x}$ and $\mathbf{x}'$ correspond with the epipoles in the two images. However, there is no problem transferring a point lying on the baseline between views two and three, or anywhere else on the trifocal plane. This is the key difference between epipolar transfer and transfer using the trifocal tensor. The former is undefined for *any* point on the trifocal plane.

15.3.3 Line transfer using the trifocal tensor

Using the trifocal tensor, it is possible to transfer lines from a pair of images to a third according to the line-transfer equation $l_i = l'_j l''_k \mathcal{T}_i^{jk}$ of table 15.2. This gives an explicit formula for the line in the first view, given lines in the other two views. Note however that if the lines $\mathbf{l}$ and $\mathbf{l}'$ are known in the first and second views then $\mathbf{l}''$ may be computed by solving the set of linear equations $(l_r\epsilon^{ris})l'_j l''_k \mathcal{T}_i^{jk} = 0^s$, thereby transferring it into the third image. Similarly one may transfer lines into the second image. Line transfer is not possible using only the fundamental matrices.

Degeneracies. Consider the geometry of figure 12.8(*p*322). The line $\mathbf{L}$ in 3-space is defined by the intersection of the planes through $\mathbf{l}$ and $\mathbf{l}'$, namely $\boldsymbol{\pi}$ and $\boldsymbol{\pi}'$ respectively. This line is clearly undefined when the planes $\boldsymbol{\pi}$ and $\boldsymbol{\pi}'$ are coincident, i.e. in the case of epipolar planes. Consequently, lines cannot be transferred between the first and third image if both $\mathbf{l}$ and $\mathbf{l}'$ are corresponding epipolar lines for the first and second views. Algebraically, the line-transfer equation gives $l_i = l'_j l''_k \mathcal{T}_i^{jk} = \mathbf{0}$, and the equation matrix $(l_r\epsilon^{ris})l'_j \mathcal{T}_i^{jk}$ used to solve for $\mathbf{l}''$ becomes zero. It is quite common for lines to be near epipolar, and their transfer is then inaccurate, so this condition should always be checked for. There is an equivalent degeneracy for line transfer between views one and two defined by a line in view three. Again, it occurs if the lines in views one and three are corresponding epipolar lines for these two views.

In general the epipolar geometries between views one and two, and one and three will differ, for instance the epipole $\mathbf{e}_{12}$ arising in the first view from view two will not coincide with the epipole $\mathbf{e}_{13}$ arising in the first view from view three. Thus an epipolar line in the first view for views one and two will not coincide with an epipolar line for views one and three. Consequently, when line transfer into the third view is degenerate, line transfer into the second view will not in general be degenerate. However, for lines in the trifocal plane transfer is degenerate (i.e. undefined) always.

15.4 The fundamental matrices for three views

The three fundamental matrices $\mathtt{F}_{21}, \mathtt{F}_{31}, \mathtt{F}_{32}$ are not independent, but satisfy three relations:

$$\mathbf{e}_{23}^{\mathsf{T}}\,\mathtt{F}_{21}\,\mathbf{e}_{13} = \mathbf{e}_{31}^{\mathsf{T}}\,\mathtt{F}_{32}\,\mathbf{e}_{21} = \mathbf{e}_{32}^{\mathsf{T}}\,\mathtt{F}_{31}\,\mathbf{e}_{12} = 0. \tag{15.11}$$

These relations are easily seen from figure 15.8. For example, $\mathbf{e}_{32}^{\mathsf{T}}\,\mathtt{F}_{31}\,\mathbf{e}_{12} = 0$ follows from the observation that $\mathbf{e}_{32}$ and $\mathbf{e}_{12}$ are matching points, corresponding to the centre of camera number 2.

Projectively, the three-camera configuration has 18 degrees of freedom counting 11 for each camera less 15 for an overall projective ambiguity. Alternatively, this may be accounted for as 21 for the 3×7 degrees of freedom of the fundamental matrices less 3 for the relations. The trifocal tensor also has 18 degrees of freedom and fundamental matrices computed from the trifocal tensor will automatically satisfy the three relations.

The counting argument implies that the three relations of (15.11) are sufficient to ensure consistency of three fundamental matrices. The counting argument alone is not a convincing proof of this, however, so a proof is given below.

Definition 15.5. Three fundamental matrices $\mathtt{F}_{21}$, $\mathtt{F}_{31}$ and $\mathtt{F}_{32}$ are said to be *compatible* if they satisfy the conditions (15.11).

In most cases, these conditions are sufficient to ensure that the three fundamental matrices correspond to some geometric configuration of cameras.

Theorem 15.6. *Let a set of three fundamental matrices* $\mathtt{F}_{21}$, $\mathtt{F}_{31}$ *and* $\mathtt{F}_{32}$ *be given satisfying the conditions (15.11). Assume also that* $\mathbf{e}_{12} \neq \mathbf{e}_{13}$, $\mathbf{e}_{21} \neq \mathbf{e}_{23}$, *and* $\mathbf{e}_{31} \neq \mathbf{e}_{32}$. *Then there exist three camera matrices* $\mathtt{P}_1, \mathtt{P}_2, \mathtt{P}_3$ *such that* $\mathtt{F}_{ij}$ *is the fundamental matrix corresponding to the pair* $(\mathtt{P}_i, \mathtt{P}_j)$.

Note that the conditions $\mathbf{e}_{ij} \neq \mathbf{e}_{ik}$ in this theorem ensure that the three cameras are non-collinear. For this reason they will be referred to here as the *non-collinearity conditions*. One may show by example (left to the reader) that these conditions are necessary for the truth of the theorem.

Proof. In this proof, the indices i, j and k are intended to be distinct. We begin by choosing three points $\mathbf{x}_i; i = 1, \ldots, 3$, consistent with the three fundamental matrices. In other words, we require that $\mathbf{x}_i^{\mathsf{T}} \mathtt{F}_{ij} \mathbf{x}_j = 0$ for all pairs (i, j). This is easily done by choosing first $\mathbf{x}_1$ and $\mathbf{x}_2$ to satisfy $\mathbf{x}_2^{\mathsf{T}} \mathtt{F}_{21} \mathbf{x}_1 = 0$, and then defining $\mathbf{x}_3$ to be the intersection of the two epipolar lines $\mathtt{F}_{32}\mathbf{x}_2$ and $\mathtt{F}_{31}\mathbf{x}_1$.

In a similar manner, we choose a second set of points $\mathbf{y}_i; i = 1, \ldots, 3$ satisfying $\mathbf{y}_i^{\mathsf{T}} \mathtt{F}_{ij} \mathbf{y}_j = 0$. This is done in such a way that the four points $\mathbf{x}_i, \mathbf{y}_i, \mathbf{e}_{ij}, \mathbf{e}_{ik}$ in each image i are in general position – that is no three are collinear. This is possible by the assumption that the two epipoles in each image are distinct.

Next we choose five world points $\mathbf{C}_1, \mathbf{C}_2, \mathbf{C}_3, \mathbf{X}, \mathbf{Y}$ in general position. For example, one could take the usual projective basis. We may now define the three camera matrices. Let the i-th camera matrix $\mathtt{P}_i$ satisfy the conditions

$$\mathtt{P}_i \mathbf{C}_i = \mathbf{0}; \quad \mathtt{P}_i \mathbf{C}_j = \mathbf{e}_{ij}; \quad \mathtt{P}_i \mathbf{C}_k = \mathbf{e}_{ik}; \quad \mathtt{P}_i \mathbf{X} = \mathbf{x}_i; \quad \mathtt{P}_i \mathbf{Y} = \mathbf{y}_i.$$

In other words, the i-th camera has centre at $\mathbf{C}_i$ and maps the four other world points $\mathbf{C}_j, \mathbf{C}_k, \mathbf{X}, \mathbf{Y}$ to the four image points $\mathbf{e}_{ij}, \mathbf{e}_{ik}, \mathbf{x}_i, \mathbf{y}_i$. This uniquely determines the camera matrix since the points are in general position. To see this, recall that the camera matrix defines a homography between the image and the rays through the camera centre (a 2D projective space). The images of four points specify this homography completely. Let $\hat{\mathtt{F}}_{ij}$ be the fundamental matrix defined by the pair of camera matrices $\mathtt{P}_i$ and $\mathtt{P}_j$. The proof is completed by proving that $\hat{\mathtt{F}}_{ij} = \mathtt{F}_{ij}$ for all i, j.

The epipoles of $\hat{\mathtt{F}}_{ij}$ and $\mathtt{F}_{ij}$ are the same, by the way that $\mathtt{P}_i$ and $\mathtt{P}_j$ are constructed. Consider the pencil of epipolar lines through $\mathbf{e}_{ij}$ in image i. This pencil forms a 1-dimensional projective space of lines, and the fundamental matrix $\mathtt{F}_{ij}$ induces a one-to-one correspondence (in fact a homography) between this pencil and the pencil of lines through $\mathbf{e}_{ji}$ in image j. The fundamental matrix $\hat{\mathtt{F}}_{ij}$ also induces a homography between the same pencils. The two fundamental matrices are the same if the homographies they induce are the same.

Two 1-dimensional homographies are the same if they agree on three points (or in this case epipolar lines). The relation $\mathbf{x}_i^{\mathsf{T}} \mathtt{F}_{ij} \mathbf{x}_j = 0$ means that the epipolar lines through $\mathbf{x}_i$ in image i and $\mathbf{x}_j$ in image j correspond under the homography induced by $\mathtt{F}_{ij}$. By construction $\mathbf{x}_i \hat{\mathtt{F}}_{ij} \mathbf{x}_j = 0$ as well, since $\mathbf{x}_i$ and $\mathbf{x}_j$ are the projections of the point $\mathbf{X}$ in the two images. Thus, both homographies agree on this pair of epipolar lines. In the same way, the homographies induced by $\mathtt{F}_{ij}$ and $\hat{\mathtt{F}}_{ij}$ agree on the epipolar lines corresponding to the pairs $\mathbf{y}_i \leftrightarrow \mathbf{y}_j$ and $\mathbf{e}_{ik} \leftrightarrow \mathbf{e}_{jk}$. The two homographies therefore agree on three lines in the pencil and hence are equal; so are the corresponding fundamental matrices. (We are grateful to Frederik Schaffalitzky for this proof). □

15.4.1 Uniqueness of camera matrices given three fundamental matrices

The proof just given shows that there is at least one set of cameras corresponding to three compatible fundamental matrices (provided they satisfying the non-collinearity condition). It is important to know that the three fundamental matrices determine the configuration of the three cameras uniquely, at least up to the unavoidable projective ambiguity. This will be shown next.

The first two camera matrices $\mathtt{P}$ and $\mathtt{P}'$ may be determined from the fundamental matrix $\mathtt{F}_{21}$ by two-view techniques (chapter 9). It remains to determine the third camera matrix $\mathtt{P}''$ in the same projective frame. In principle, this may be done as follows.

(i) Select a set of matching points $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ in the first two images, satisfying $\mathbf{x}_i'^{\mathsf{T}}\mathtt{F}_{21}\mathbf{x}_i = 0$, and use triangulation to determine the corresponding 3D points $\mathbf{X}_i$.

(ii) Use epipolar transfer to determine the corresponding points $\mathbf{x}''_i$ in the third image, using the fundamental matrices $\mathtt{F}_{31}$ and $\mathtt{F}_{32}$.

(iii) Solve for the camera matrix $\mathtt{P}''$ from the set of 3D–2D correspondences $\mathbf{X}_i \leftrightarrow \mathbf{x}''_i$.

The second step in this algorithm will fail in the case where the point $\mathbf{X}_i$ lies in the trifocal plane. Such a point $\mathbf{X}_i$ is easily detected and discarded, since it projects into the first image as a point $\mathbf{x}_i$ lying on the line joining the two epipoles $\mathbf{e}_{12}$ and $\mathbf{e}_{13}$. Since there are infinitely many possible matched points, we can compute sufficiently many such points to compute $\mathtt{P}''$.

The only situation in which this method will fail is when all space points $\mathbf{X}_i$ lie in a trifocal plane. This can occur only in the degenerate situation in which the three camera centres are collinear, in which case the trifocal plane is not uniquely determined. Thus, we see that unless the three camera centres are collinear, the three camera matrices may be determined from the fundamental matrices. On the other hand, if the three cameras are collinear, then there is no way to determine the relative spacings of the cameras along the line of their centres. This is because the length of the baseline cannot be determined from the fundamental matrices, and the three baselines (distances between the camera centres) may be arbitrarily chosen and remain consistent with the fundamental matrices. Thus we have demonstrated the following fact:

Result 15.7. *Given three compatible fundamental matrices $\mathtt{F}_{21}$, $\mathtt{F}_{31}$ and $\mathtt{F}_{32}$ satisfying the non-collinearity condition, the three corresponding camera matrices $\mathtt{P}$, $\mathtt{P}'$ and $\mathtt{P}''$ are unique up to the choice of a 3D projective coordinate frame.*

15.4.2 Computation of camera matrices from three fundamental matrices

Given three compatible fundamental matrices, there exists a simple method for computing a corresponding set of three camera matrices. From the fundamental matrix $\mathtt{F}_{21}$, one can compute a corresponding pair of camera matrices $(\mathtt{P}, \mathtt{P}')$ using result 9.14-(*p*256). Next, according to result 9.12(*p*255) the third camera matrix $\mathtt{P}''$ must satisfy the condition that $\mathtt{P}''^{\mathsf{T}}\mathtt{F}_{31}\mathtt{P}$ and $\mathtt{P}''^{\mathsf{T}}\mathtt{F}_{32}\mathtt{P}'$ be skew-symmetric. Each of these matrices gives rise to 10 linear equations in the entries of $\mathtt{P}''$, a total of 20 equations in the 12 entries of $\mathtt{P}''$. From these, $\mathtt{P}''$ may be computed linearly.

If the three fundamental matrices are compatible in the sense of definition 15.5 and the non-collinearity condition of theorem 15.6 holds, then there will exist a solution, and it will be unique. If however the three fundamental matrices are computed independently from point correspondences, then they will not satisfy the compatibility conditions exactly. In this case it will be necessary to compute a least-squares solution to find $\mathtt{P}''$. The error being minimized is not geometrically based. It is best to use this algorithm only when the fundamental matrices are known to be compatible.

One can think of doing three-view reconstruction by estimating the three fundamental matrices using pairwise point correspondences, then using the above algorithm to estimate the three camera matrices. This is not a very good strategy, for the following reasons.

(i) The method for computing the three camera matrices from the fundamental matrices assumes that the fundamental matrices are compatible. Otherwise, a least-squares problem involving a non-geometrically justified cost function is involved.

(ii) Although result 15.7 shows that three fundamental matrices may determine the camera geometry, and hence the trifocal tensor, this is only true when the cameras are not collinear. As they approach collinearity, the estimate of the relative camera placement becomes unstable.

The trifocal tensor is preferable to a triple of compatible fundamental matrices as a means of determining the geometry of three views. This is because the difficulty with the views being collinear is not an issue with the trifocal tensor. It is well defined and uniquely determines the geometry even for collinear cameras. The difference is that the fundamental matrices do not contain a direct constraint on the relative displacements between the three cameras, whereas this is built into the trifocal tensor.

Since the projective structure of the three cameras may be computed explicitly from the trifocal tensor, it follows that all three fundamental matrices for the three view pairs are determined by the trifocal tensor. In fact simple formulae, given in algorithm 15.1-(*p*375) exist for the two fundamental matrices $\mathtt{F}_{21}$ and $\mathtt{F}_{31}$. The fundamental matrices determined from the trifocal tensor will satisfy the compatibility conditions (15.11).

15.4.3 Camera matrices compatible with two fundamental matrices

Suppose we are given only two fundamental matrices $\mathtt{F}_{21}$ and $\mathtt{F}_{31}$. To what extent do these fix the geometry of the three cameras? It will be shown here that there are four degrees of freedom in the solution for the camera matrices, beyond the usual projective ambiguity.

From $\mathtt{F}_{21}$ one may compute a pair of camera matrices $(\mathtt{P}, \mathtt{P}')$, and from $\mathtt{F}_{31}$ the pair $(\mathtt{P}, \mathtt{P}'')$. In both cases we may choose $\mathtt{P} = [\mathtt{I} \mid \mathbf{0}]$, resulting in a triple of camera matrices $(\mathtt{P}, \mathtt{P}', \mathtt{P}'')$ compatible with the pair of fundamental matrices.

However, the choice of the three camera matrices is not unique, since for any matrices $\mathtt{H}_1$ and $\mathtt{H}_2$ representing 3D projective transforms, the pairs $(\mathtt{PH}_1, \mathtt{P}'\mathtt{H}_1)$ and $(\mathtt{PH}_2, \mathtt{P}''\mathtt{H}_2)$ are also compatible with the same fundamental matrices. In order to preserve the condition that $\mathtt{P}$ is equal to $[\mathtt{I} \mid \mathbf{0}]$ in each case, the form of $\mathtt{H}_i$ must be restricted to:

$$\mathtt{H}_i = \begin{bmatrix} \mathtt{I} & \mathbf{0} \\ \mathbf{v}_i^{\mathsf{T}} & k_i \end{bmatrix}.$$

We may now fix on a particular choice of the first two camera matrices $(\mathtt{P}, \mathtt{P}')$ compatible with $\mathtt{F}_{21}$. This is equivalent to fixing on a specific projective coordinate frame. The general solution for the camera matrices is then $(\mathtt{P}, \mathtt{P}', \mathtt{P}''\mathtt{H}_2)$, where $\mathtt{H}_2$ is of the form given above and the two pairs $(\mathtt{P}, \mathtt{P}')$ and $(\mathtt{P}, \mathtt{P}'')$ are compatible with the two fundamental matrices.

Allowing also for the overall projective ambiguity, the most general solution is $(\mathtt{PH}, \mathtt{P}'\mathtt{H}, \mathtt{P}''\mathtt{H}_2\mathtt{H})$, which gives a total of 19 degrees of freedom, 15 for the projective transformation $\mathtt{H}$ and 4 for the degrees of freedom of $\mathtt{H}_2$. The same number of degrees of freedom may be found using a counting argument as follows: two fundamental matrices have 7 degrees of freedom each, for a total of 14. Three arbitrary camera matrices on the other hand have $3 \times 11 = 33$ degrees of freedom. The 14 constraints imposed by the two fundamental matrices leave 19 remaining degrees of freedom for the three camera matrices.

15.5 Closure

The development of three-view geometry proceeds in an analogous manner to that of two-view geometry covered in part II of this book. The trifocal tensor may be computed from image correspondences over three views, and a projective reconstruction of the cameras and 3D scene then follows. This computation is described in chapter 16. The projective ambiguity may be reduced to affine or metric by supplying additional information on the scene or cameras in the same manner as that of chapter 10. A similar development to that of chapter 13 may be given for the relations between homographies induced by scene planes and the trifocal tensor.

15.5.1 The literature

With hindsight, the discovery of the trifocal tensor may be traced to [Spetsakis-91] and [Weng-88], where it was used for scene reconstruction from lines in the case of calibrated cameras. It was later shown in [Hartley-94d] to be equally applicable to projective scene reconstruction in the uncalibrated case. At this stage matrix notation was used, but [Vieville-93] used tensor notation for this problem.

Meanwhile in independent work, Shashua introduced trilinearity conditions relating the coordinates of corresponding points in three views with uncalibrated cameras [Shashua-94, Shashua-95a]. [Hartley-95b, Hartley-97a] then showed that Shashua's relation for points and scene reconstruction from lines both arise from a common tensor, and the trifocal tensor was explicitly identified.

In subsequent work properties of the tensor have been investigated, e.g. [Shashua-95b]. In particular [Triggs-95] described the mixed covariant–contravariant behaviour of the indices, and [Zisserman-96] described the geometry of the homographies encoded by the tensor. Faugeras and Mourrain [Faugeras-95a] gave enlightening new derivations of the trifocal tensor equations and considered the trifocal tensor in the context of general linear constraints involving multiple views. This approach will be discussed in chapter 17. Further geometric properties of the tensor were given in Faugeras & Papadopoulo [Faugeras-97].

Epipolar point transfer was described by [Barrett-92, Faugeras-94], and its deficiencies pointed out by [Zisserman-94], amongst others.

The trifocal tensor has been used for various applications including establishing correspondences in image sequences [Beardsley-96], independent motion detection [Torr-95a], and camera self-calibration [Armstrong-96a].

15.5.2 Notes and exercises

(i) The trifocal tensor is invariant to 3D projective transforms. Verify explicitly that if $\mathtt{H}_{4\times 4}$ is a transform preserving the first camera matrix $\mathtt{P} = [\mathtt{I} \mid \mathbf{0}]$, then the tensor defined by (15.1–*p*367) is unchanged.

(ii) In this chapter the starting point for the trifocal tensor derivation was the incidence property of three corresponding lines. Show that alternatively the starting point may be the homography induced by a plane.

Here is a sketch derivation: choose the camera matrices to be a canonical set $\mathtt{P} = [\mathtt{I} \mid \mathbf{0}], \mathtt{P}' = [\mathtt{A} \mid \mathbf{a}_4]$, $\mathtt{P}'' = [\mathtt{B} \mid \mathbf{b}_4]$ and start from the homography $\mathtt{H}_{13}$ between the first and third views induced by a plane $\boldsymbol{\pi}'$. From result 13.1(*p*326) this homography may be written as $\mathtt{H}_{13} = \mathtt{B} - \mathbf{b}_4\mathbf{v}^\mathsf{T}$, where $\boldsymbol{\pi}'^\mathsf{T} = (\mathbf{v}^\mathsf{T}, 1)$. In this case the plane is defined by a line $\mathbf{l}'$ in the second view as $\boldsymbol{\pi}' = \mathtt{P}'^\mathsf{T}\mathbf{l}'$. Show that result 15.2(*p*369) follows.

(iii) Homographies involving the first view are simply expressed in terms of the trifocal tensor $\mathcal{T}_i^{jk}$ as given by result 15.2(*p*369). Investigate whether a simple formula exists for the homography $\mathtt{H}_{23}$ from the second to the third view, induced by a line $\mathbf{l}$ in the first image.

(iv) The contraction $x^i\mathcal{T}_i^{jk}$ is a 3×3 matrix. Show that this may be interpreted as a correlation (see definition 2.29(*p*59)) mapping between the second and third views induced by the line which is the back-projection of the point $\mathbf{x}$ in the first view.

(v) **Plane plus parallax over three views.** There is a rich geometry associated with the plane plus two points configuration (see figure 13.9(*p*336)) over three views: suppose the points off the (reference) plane are $\mathbf{X}$ and $\mathbf{Y}$. Project the point $\mathbf{X}$ onto the reference plane from each of the three camera centres to form a triangle $\mathbf{x}, \mathbf{x}', \mathbf{x}''$, and similarly project the point $\mathbf{Y}$ to the triangle $\mathbf{y}, \mathbf{y}', \mathbf{y}''$. Then the two triangles form a Desargues's configuration and are related by a planar homology (see section A7.2(*p*629)). A simple sketch shows that the lines joining corresponding triangle vertices, $(\mathbf{x}, \mathbf{y}), (\mathbf{x}', \mathbf{y}'), (\mathbf{x}'', \mathbf{y}'')$, are concurrent, and their intersection is the point at which the line joining $\mathbf{X}$ and $\mathbf{Y}$ pierces the reference plane. Similarly the intersection points of corresponding triangle sides are collinear, and the line so formed is the intersection of the trifocal plane of the cameras with the reference plane. Further details are given in [Criminisi-98, Irani-98, Triggs-00b].

(vi) In the case where two of the three cameras have the same camera centre, the trifocal tensor may be related to simpler entities. There are two cases.

   (a) If the second and third camera have the same centre, then $\mathcal{T}_i^{jk} = \mathtt{F}_{ri}\mathtt{H}_s^k\epsilon^{rjs}$, where $\mathtt{F}_{ri}$ is the fundamental matrix for the first two views, and $\mathtt{H}$ is the homography from the second to the third view induced by the fact that they have the same centre.
   (b) If the first and the second views have the same centre, then $\mathcal{T}_i^{jk} = \mathtt{H}_i^j\mathrm{e}''^k$, where $\mathtt{H}$ is the homography from the first to the second view and $\mathbf{e}''$ is the epipole in the third image.

Prove these relationships using the approach of chapter 17.

(vii) Consider the case of a small baseline between the cameras and derive a differential form of the trifocal tensor, see [Astrom-98, Triggs-99b].

(viii) There are actually three different trifocal tensors relating three views, depending on which of the three cameras corresponds to the covariant index. Given one such tensor $[\mathtt{T}_i]$, verify that the tensor $[\mathtt{T}'_i]$ may be computed in several steps, as follows:

   (a) Extract the three camera matrices $\mathtt{P} = [\mathtt{I} \mid \mathbf{0}]$, $\mathtt{P}'$ and $\mathtt{P}''$ from the trifocal tensor.
   (b) Find a 3D projective transformation $\mathtt{H}$ such that $\mathtt{P}'\mathtt{H} = [\mathtt{I} \mid \mathbf{0}]$, and apply it to each of $\mathtt{P}$ and $\mathtt{P}''$ as well.
   (c) Compute the tensor $[\mathtt{T}'_i]$ by applying (15.1–*p*367).

(ix) Investigate the form and properties (e.g. rank of the matrices $\mathtt{T}_i$) of the trifocal tensor for the special motions (pure translation, planar motion) described in section 9.3(*p*247) for the fundamental matrix.

(x) Comparison of the incidence relationships of table 15.3(*p*378) indicates that one may replace a line l'_j by the expression $\epsilon_{jrs}x'_r$, and proceed similarly with l''_k. Also, one gets a three-view line equation by replacing x^i by $\epsilon^{irs}l_i$. Can both of these operations be carried out at once to obtain an equation

$$\left(\epsilon^{iru}l_i\right)\left(\epsilon_{jsv}x'^j\right)\left(\epsilon_{ktw}x''^k\right)\mathcal{T}_r^{st} = 0^u_{vw}?$$

Why, or why not?

(xi) **Affine trifocal tensor.** If the three cameras $\mathtt{P}, \mathtt{P}'$ and $\mathtt{P}''$ are all affine (definition 6.3(*p*166)), then the corresponding tensor $\mathcal{T}_{\mathrm{A}}$ is the *affine trifocal tensor*. This affine specialization of the tensor has 12 degrees of freedom and 16 non-zero entries. The affine trifocal tensor was first defined in [Torr-95b], and has been studied in [Kahl-98a, Quan-97a, Thorhallsson-99]. It shares with the affine fundamental matrix (chapter 14) very stable numerical estimation behaviour. It has been shown to perform very well in tracking applications where the object of interest (for example a car) has small relief compared to the depth of the scene [Hayman-03, Tordoff-01].