CHAPTER 33

Pairs of Cameras

33.1 GEOMETRY

Place two perspective cameras in 3D, and construct the line segment joining their focal points. This line segment is known as the *baseline*. Now extend the baseline to a line. This line intersects each image plane in an important point, known as the *epipole* for that image plane in that configuration of cameras (Figure 33.1).

If the cameras view some point $\mathbf{X}$ into 3D, camera 1 sees that point at $\mathbf{x}_1$ and camera 2 sees that point at $\mathbf{x}_2$. The points $\mathbf{X}$, $\mathbf{f}_1$ and $\mathbf{f}_2$ define a plane in 3D. This plane intersects the image plane of camera 1 in a line I shall call $\mathbf{l}_1(\mathbf{X})$. Similarly, it intersects camera 2 in an line $\mathbf{l}_2(\mathbf{X})$. These lines must pass through their respective epipoles (Figure 33.2), and are known as *epipolar lines*.

As section 22.6 sketched, knowing something about the relative geometry of the cameras and where the point appears in each camera will reveal the 3D coordinates of the point. If you see a point in the first camera at $\mathbf{x}_1$, you will need to find $\mathbf{x}_2$ to produce a 3D reconstruction. But not any point in camera 2 could correspond to $\mathbf{x}_1$. If you know enough about relative camera configuration, you can construct the epipolar line $\mathbf{l}_2$ in the second image, and $\mathbf{x}_2$ must lie on this line.

The epipoles contain further valuable information. Introduce a second point $\mathbf{Y}$ that doesn't lie on the plane through $\mathbf{X}$, $\mathbf{f}_1$ and $\mathbf{f}_2$. This yields a second plane, which produces its own epipolar lines (Figure 33.3). The baseline defines a whole family of planes that contain the baseline (Figure 33.4; this is often referred to as a *star* of planes).

The epipole in a single image reveals a great deal of information about the camera has moved. For intuition, construct some of the lines through the epipole, then compare to Figure 33.4. Figure 33.5 shows some examples.

33.1.1 The Fundamental Matrix

As Figure **??** shows, a point in 3D selects a plane from the family of planes through both focal points, and this plane intersects each image plane in epipolar lines. The image of that 3D point in camera 1 selects the same epipolar plane, so the figure illustrates a mapping from points in camera 1 to lines through the epipole in camera 2, which works like this: Select a point in camera 1; construct the plane through this point and the two focal points; now intersect that plane with camera 2's image plane; and you have the corresponding epipolar line.

The point $\mathbf{x}_1$ in Figure **??** is a point in 3D (though lying on the camera plane) and can be written with four homogenous coordinates. Write $\mathbf{P}$ for the 3D coordinates of an arbitrary point on the plane through $\mathbf{x}_1$, $\mathbf{f}_1$ and $\mathbf{f}_2$. Then

$$\text{determinant}([\mathbf{x}_1, \mathbf{f}_1, \mathbf{f}_2, \mathbf{P}]) = 0$$

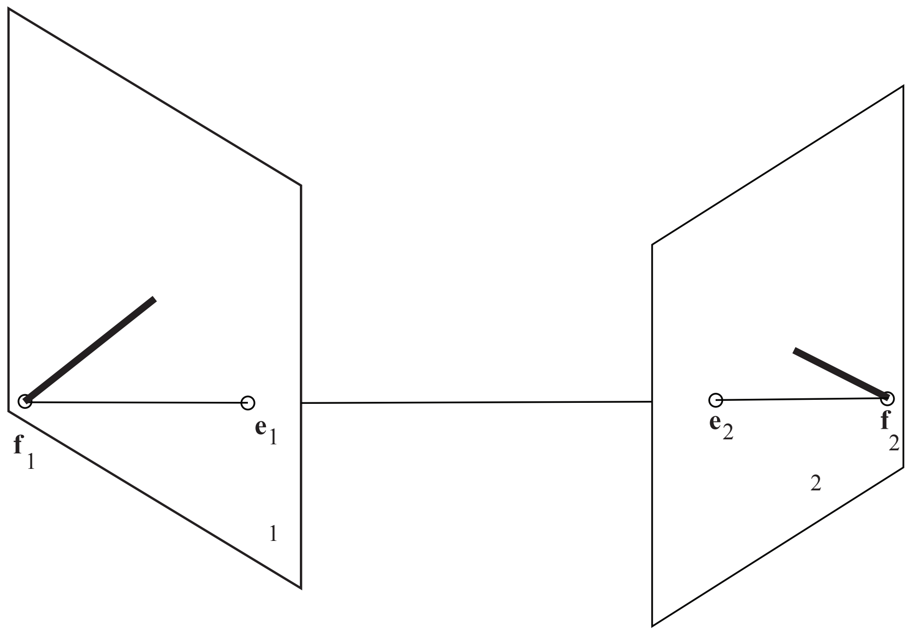

FIGURE 33.1: *Epipoles are formed by the line connecting the focal points ($\mathbf{f}_1$ and $\mathbf{f}_2$) of two perspective cameras. This line intersects camera 1's image plane in $\mathbf{e}_1$ (the epipole in the first image), and camera 2's image plane in $\mathbf{e}_2$ (the epipole in the second image). As long as the two focal points are distinct, the epipoles are properly defined, although they may appear far outside the image (or even at infinity).*

(the four points are coplanar, **exercises**). You could write this plane as $pP_1 + qP_2 + rP_3 + sP_4 = 0$, where p, q, r and s are linear functions of $\mathbf{x}_1$. The line in camera 2's image plane is obtained by intersecting this plane with a fixed plane. This means in turn that the coefficients of this line are also linear functions of $\mathbf{x}_1$. Now notice that you could write $\mathbf{x}_1$ with only three homogenous coordinates – the fourth follows from the fact that this point lies on a fixed, known plane. From these observations, it follows that there is a matrix $\mathcal{F}$ with the property that the coefficients of the epipolar line corresponding to $\mathbf{x}_1$ are given by $\mathbf{x}_1^T\mathcal{F}$. This matrix is known as the *fundamental matrix*. Here $\mathbf{x}_1^T\mathcal{F}\mathbf{x}_2' = 0$, because $\mathbf{x}_2'$ lies on this line.

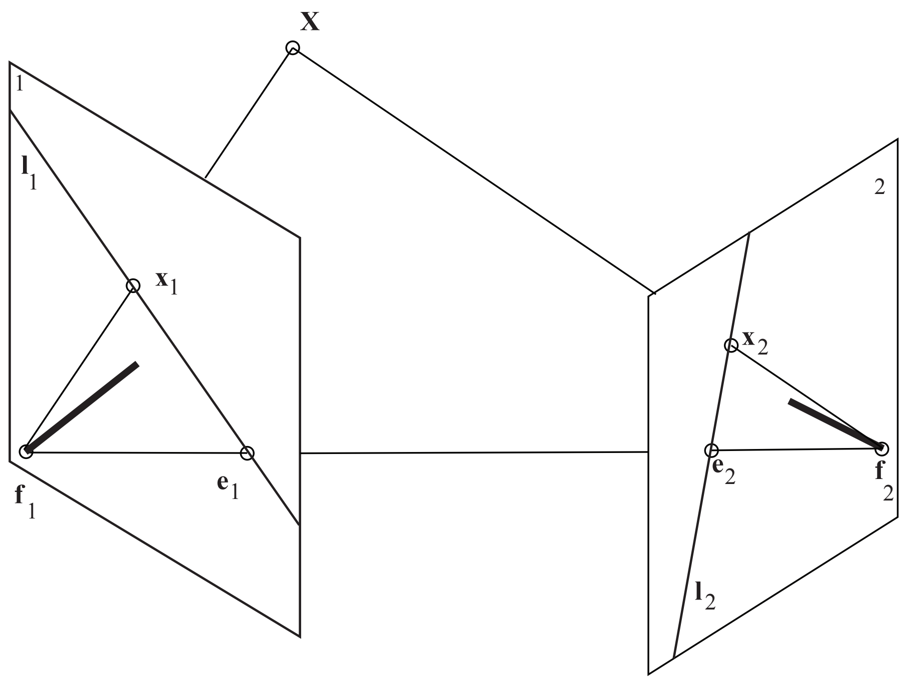

FIGURE 33.2: *The cameras are now viewing a point* $\mathbf{X}$ *in 3D, which projects to* $\mathbf{x}_1$ *in camera* 1 *and* $\mathbf{x}_2$ *in camera* 2. *The three points* $\mathbf{f}_1$, $\mathbf{f}_2$ *and* $\mathbf{X}$ *define a plane in 3D. This plane intersects camera* 1*'s image plane in a line (*$\mathbf{l}_1$ *in the figure) that passes through* $\mathbf{e}_1$ *and camera* 2*'s image plane in a line (*$\mathbf{l}_2$ *in the figure) that passes through* $\mathbf{e}_2$.

Remember this: *For any pair of cameras which do not share a focal point, there is a fundamental matrix* $\mathcal{F}$ *with the property that for any pair* $\mathbf{x}_1$, $\mathbf{x}'_2$, *where* $\mathbf{x}_1$ *is the image of a 3D point in the first camera and* $\mathbf{x}'_2$ *is the image of that point in the second camera,*

$$\mathbf{x}_1^T \mathcal{F} \mathbf{x}'_2 = 0$$

33.1.2 Properties of the Fundamental Matrix

There isn't any particular reason the cameras are labelled 1 and 2 – you could swap the labels, without affecting the geometry. Notice that $\mathbf{x}_1^T \mathcal{F} \mathbf{x}'_2 = 0$ implies that ${\mathbf{x}'_2}^T \mathcal{F}^T \mathbf{x}_1 = 0$. This means that, if you swap the labels, you transpose the fundamental matrix. In turn, a procedure that finds something in camera 2 using data from camera 1 can also be used to find something in camera 1 using data from

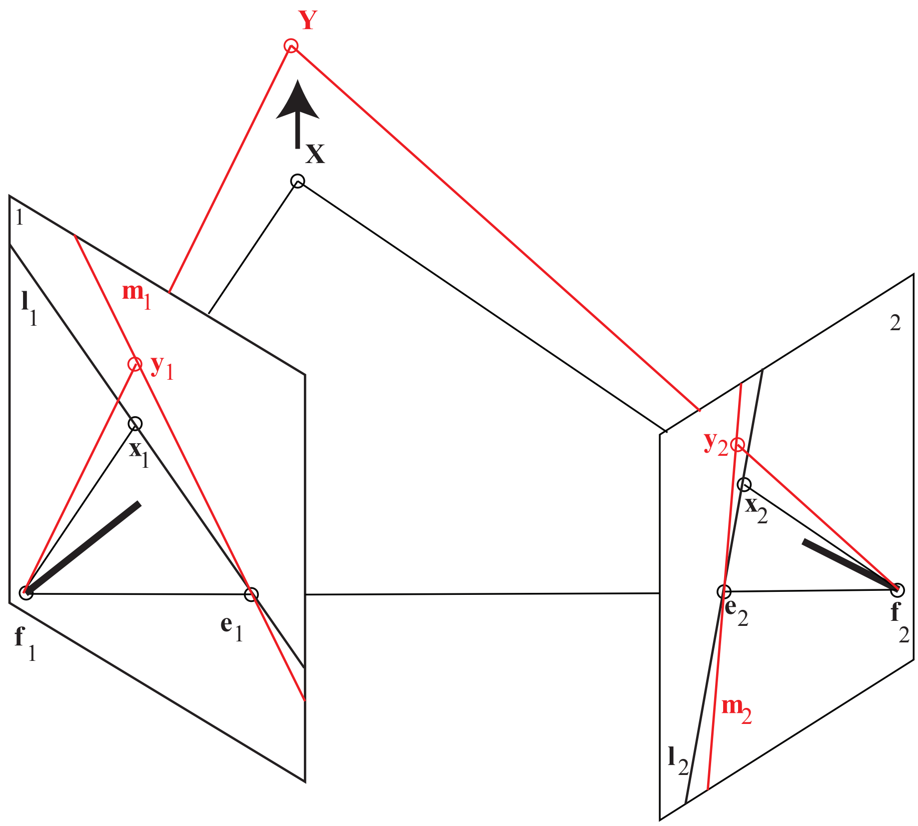

FIGURE 33.3: *The cameras now view two points* $\mathbf{X}$ *and* $\mathbf{Y}$ *in 3D. Each of these points defines a plane when taken together with the two focal points. Depending on where* $\mathbf{Y}$ *is, the two planes may be different (as in this figure). However, they both intersect each camera's image plane in lines, and these lines pass through the epipoles.*

camera 2.

The fundamental matrix for a pair of cameras reveals both epipoles and epipolar lines for both cameras. Choose some point $\mathbf{x}_1$ in the first camera. Now for *every* point $\mathbf{x}'$ in camera 2 that could match $\mathbf{x}_1$,

$$\mathbf{x}_1^T \mathcal{F} \mathbf{x}' = \left(\mathbf{x}_1^T \mathcal{F}\right) \mathbf{x}' = 0$$

and you can think of $\mathcal{F}^T \mathbf{x}_1$ as a vector containing the coefficients of a line. This line is the epipolar line corresponding to $\mathbf{x}_1$. You can think of the fundamental matrix as a map from points in one camera to lines in the other camera.

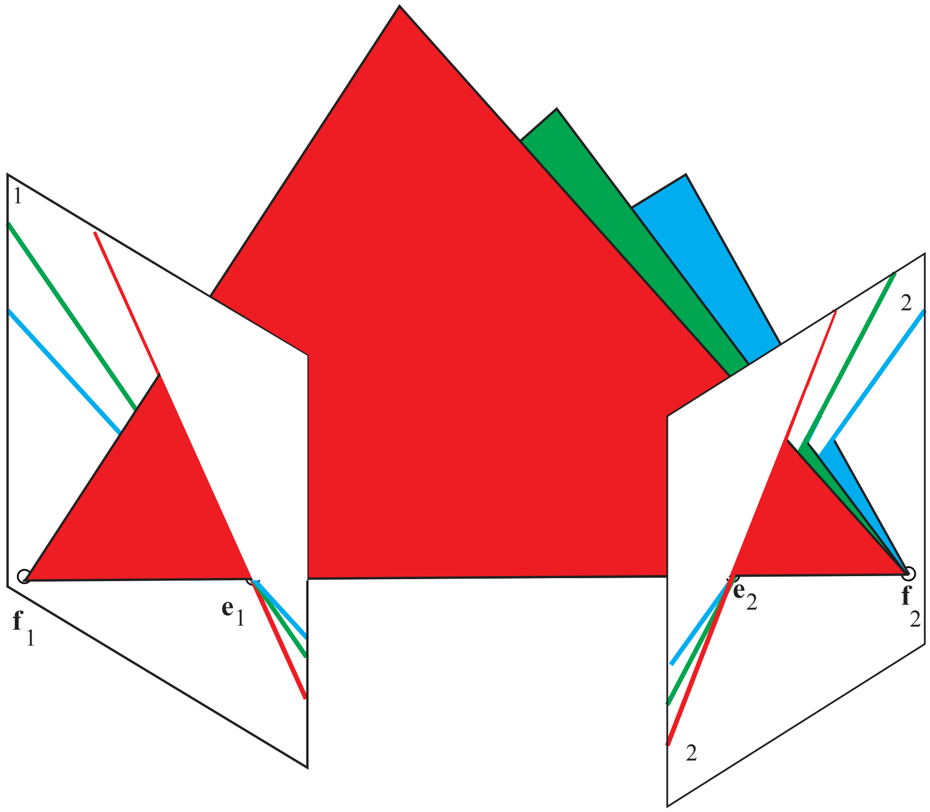

FIGURE 33.4: *There is a family of planes passing through the baseline. Each plane intersects an image plane in a line through the epipole. Alternatively, you can see the epipole in each image as defining a family of lines that pass through the epipole.*

Procedure: 33.1 *Obtaining an epipolar line from a fundamental matrix*

The epipolar line in camera 2 corresponding to $\mathbf{x}_1$ in camera 1 consists of the set of points $\mathbf{x}'$ in camera 2 which satisfy the equation

$$\left(\mathbf{x}_1^T \mathcal{F}\right) \mathbf{x}' = 0$$

and the coefficients of the line are

$$\left(\mathcal{F}^T \mathbf{x}_1\right).$$

It follows that the coefficients of the epipolar line in camera 1 corresponding to $\mathbf{x}_2'$ in camera 2 are

$$(\mathcal{F}\mathbf{x}_2').$$

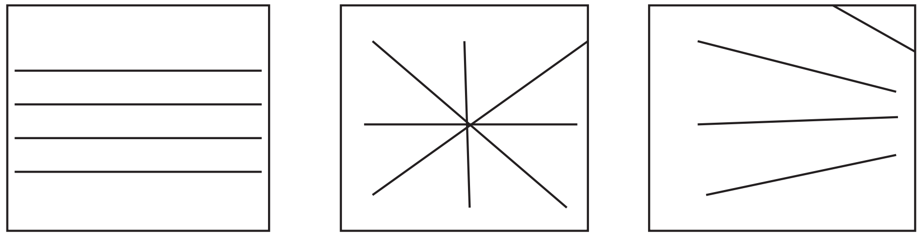

FIGURE 33.5: *The epipoles reveal information about how cameras have moved. This is most easily seen by thinking about the epipolar lines. These three examples sketch the underlying intuition, and should be looked at together with Figure 33.4. On the* **left**, *the camera has translated parallel to the image plane;* **center**, *the camera has moved perpendicular to the image plane; and* **right**, *the camera has translated parallel to the image plane then rotated slightly. You should check each case carefully.*

Every such line passes through the epipole, so the epipole must be the point in camera 2, $\mathbf{e}_2'$, such that for *any* choice of $\mathbf{x}_1$, $\left(\mathbf{x}_1^T\mathcal{F}\right)\mathbf{e}_2' = 0$. The only way to achieve this is if

$$\mathcal{F}\mathbf{e}_2' = \mathbf{0}$$

so $\mathcal{F}$ cannot have full rank.

Procedure: 33.2 *Obtaining epipoles from a fundamental matrix*

The epipole in camera 2 is the point $\mathbf{e}_2'$ such that

$$\mathcal{F}\mathbf{e}_2' = \mathbf{0}.$$

It follows that the epipole in camera 1 is the point $\mathbf{e}_1$ such that

$$\mathcal{F}^T\mathbf{e}_1 = \mathbf{0}.$$

In fact, $\mathcal{F}$ must have rank 2. The rank can't be three, because there are epipoles. The fundamental matrix is a map from points in one camera to lines in the other. If the rank were 1, the fundamental matrix would map any point in one camera to the same line in the other camera (**exercises**). If the rank were 0, the fundamental matrix would map any point in one camera to a zero vector (this happens if the camera is not translated, a rather special case, Section **??**). Rank 2 is the only available alternative.

Remember this: *The fundamental matrix $\mathcal{F}$ must have rank 2 unless the two cameras share a focal point, when it consists of zeros.*

Since

$$\mathbf{x}_1^T \mathcal{F} \mathbf{x}_2' = 0 = \mathbf{x}_1^T (s\mathcal{F}) \mathbf{x}_2'$$

for any $s \neq 0$, the fundamental matrix is only really meaningful up to scale.

> **Remember this:** *The fundamental matrix $\mathcal{F}$ is only meaningful up to scale. You should think of this matrix as a point in 8 dimensional space represented with 9 homogenous coordinates.*

33.1.3 Estimating the Fundamental Matrix

The fundamental matrix can be estimated from point correspondences. Assume two cameras view a set of points $\mathbf{X}_i$ in 3D. Write $\mathbf{x}_{1,i}$ for the image of the i'th point in camera 1 and $\mathbf{x}'_{2,i}$ for the image of the i'th point in camera 2. Each pair yields one equation that constrains the fundamental matrix, that is

$$\mathbf{x}_{1,i}^T \mathcal{F} \mathbf{x}'_{2,i} = 0$$

(remember – you *know* the coordinates of the point in each camera, so the unknowns here are the elements of $\mathcal{F}$). You get one equation for each pair of points, and each equation is homogenous, so eight pairs of points yield an estimate of $\mathcal{F}$. This estimate is up to scale, but the fundamental matrix is only meaningful up to scale. There is a useful improvement available. The scale of the image coordinate system can have a real effect on the estimate of $\mathcal{F}$, and a more accurate estimate of $\mathcal{F}$ can be obtained by scaling the image so that the largest coordinate direction runs from 0 to 1 (rather than, say, 0 to 768). The procedure is known as the *8 point algorithm*.

> **Procedure: 33.3** *The 8 point algorithm for estimating the fundamental matrix*
>
> Scale the image coordinate system for camera 1 and camera 2 so that the largest coordinate direction runs from 0 to 1. Find 8 pairs of corresponding points $\mathbf{x}_{1,i}$ and $\mathbf{x}'_{2,i}$. Now solve the system of 8 homogenous equations given by
>
> $$\mathbf{x}_{1,i}^T \mathcal{F} \mathbf{x}'_{2,i} = 0$$
>
> in $\mathcal{F}$.

The eight point algorithm does not impose the constraint the fundamental matrix has rank 2. This constraint is cubic in the coefficients of $\mathcal{F}$ (the constraint is $\det(\mathcal{F}) = 0$). Exploiting the constraint makes it possible to estimate a fundamental matrix with seven corresponding pairs, *if* you are willing to form the roots of a cubic. This isn't for everyone; details in [].

The eight point algorithm requires 8 corresponding pairs. The natural source of these pairs is RANSAC, but notice that this is a large number of pairs, so you really do not want to just select from all pairs of points. Instead, find and describe interest points using (for example) the methods of Chapter **??**, and use only pairs whose descriptors match well. Again, this isn't for everyone; details in [].

33.2 COORDINATE GEOMETRY

The drawings of the previous section illustrate geometric facts that do not depend on coordinates, but usually you need to use two images to reconstruct a point in 3D. This problem has to be worked in coordinates. Assume each camera has known intrinsics. This means you can calibrate each camera so that the coordinates the camera reports are the coordinates of a point in the camera's coordinate system and each has focal length 1.

33.2.1 Triangulating a Point in 3D from Two Images

Now choose a left-handed coordinate system so that the first camera has focal point at the origin, looks down the z-axis, and has image plane at $z = 1$, as in Section 31.2. To get the second camera, rotate the first camera by $\mathcal{R}^T$, then translate it by $\mathbf{t}$, so that $\mathbf{f}_2 = \mathbf{t}$. Notice that this means that a point at $\mathbf{X} = [X_1, X_2, X_3]^T$ in the first camera's coordinate system appears at $\mathbf{X}' = \mathcal{R}(\mathbf{X} - \mathbf{t})$ (if the camera rotates left, then all the points in the image frame move right).

The first camera has focal point at the origin, so $\mathbf{v}_1 = \mathbf{X}$ is the vector from $\mathbf{f}_1$ to $\mathbf{X}$. The vector from $\mathbf{f}_2$ to $\mathbf{X}$ is $\mathbf{v}_2 = \mathbf{X} - \mathbf{t}$ *in the first camera's coordinate system.* I will write vectors in the second camera's coordinate system with a prime, and I will work this problem in affine coordinates. Write

$$\mathcal{R} = \begin{bmatrix} \mathbf{r}_1^T \\ \mathbf{r}_2^T \\ \mathbf{r}_3^T \end{bmatrix}$$

You see the point in the first camera at

$$\mathbf{m} = \begin{bmatrix} m_1 \\ m_2 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{X_1}{X_3} \\ \frac{X_2}{X_3} \\ 1 \end{bmatrix}$$

and in the second camera at

$$\mathbf{m}' = \begin{bmatrix} m_1' \\ m_2' \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{\mathbf{r}_1^T(\mathbf{X}-\mathbf{t})}{\mathbf{r}_3^T(\mathbf{X}-\mathbf{t})} \\ \frac{\mathbf{r}_2^T(\mathbf{X}-\mathbf{t})}{\mathbf{r}_3^T(\mathbf{X}-\mathbf{t})} \\ 1 \end{bmatrix}.$$

For the moment, assume that $\mathbf{m}$, $\mathbf{m}'$, the rotation and the translation are all known exactly. Then

$$X_3\mathbf{m} = \mathbf{X}$$

and the only unknown is X_3. You can write two linear equations in this unknown, which are

$$\begin{array}{rcl} m'_1(\mathbf{r}_3^T(X_3\mathbf{m}-\mathbf{t}))-(\mathbf{r}_1^T(X_3\mathbf{m}-\mathbf{t})) & = & 0 \\ m'_2(\mathbf{r}_3^T(X_3\mathbf{m}-\mathbf{t}))-(\mathbf{r}_2^T(X_3\mathbf{m}-\mathbf{t})) & = & 0. \end{array}$$

These equations have to be consistent, which means that there is a relationship between $\mathbf{m}$ and $\mathbf{m}'$ that depends on $\mathcal{R}$ and $\mathbf{t}$. The relationship expresses the mapping from points to lines of the previous section. It could be obtained by some aggressive linear algebra, but is better constructed directly, which I do in the next section.

A warning: it is not a good idea to *estimate* X_3 using the equations above, because you will never actually know $\mathbf{m}$ and $\mathbf{m}'$ exactly. They are useful only to establish that you can recover X_3.

33.2.2 Triangulation by Minimization

Now assume you know $\mathcal{R}$ and $\mathbf{t}$, and have estimated locations $\mathbf{m}$ and $\mathbf{m}'$ for a pair of points that correspond. These estimates may not be exact – for example, they might come from an interest point matcher – but any error is small. You must recover the point in 3D.

The first camera has camera matrix $\mathcal{C}_p$. The second camera has camera matrix

$$[\mathcal{R}|-\mathcal{R}\mathbf{t}]$$

(recall notation from Section 22.6, and check that this camera has focal point at $\mathbf{t}$). Now write $\mathbf{X}=[X_1,X_2,X_3]$ for a point in 3D in affine coordinates. The residual vector in camera 1 is the vector from the projection of $\mathbf{X}$ to $\mathbf{m}$, so

$$\mathbf{e}_1(\mathbf{X})=\left[\begin{array}{c} \frac{X_1}{X_3}-m_1 \\ \frac{X_2}{X_3}-m_2 \end{array}\right].$$

The residual vector in camera 2 is the vector from the projection of $\mathbf{X}$ to $\mathbf{m}$, so

$$\mathbf{e}_2(\mathbf{X})=\left[\begin{array}{c} \frac{\mathbf{r}_1^T(\mathbf{X}-\mathbf{t})}{\mathbf{r}_3^T(\mathbf{X}-\mathbf{t})}-m'_1 \\ \frac{\mathbf{r}_2^T(\mathbf{X}-\mathbf{t})}{\mathbf{r}_3^T(\mathbf{X}-\mathbf{t})}-m'_2 \end{array}\right].$$

The *reprojection error* $E_r(\mathbf{X})$ for a point $\mathbf{X}$ in 3D is the sum of distances in each camera from the projections of the point to the measured locations, so

$$E_r(\mathbf{X})=\mathbf{e}_1^T\mathbf{e}_1+\mathbf{e}_2^T\mathbf{e}_2.$$

It is natural to obtain $\mathbf{X}$ by simply minimizing the reprojection error.

Procedure: 33.4 *Triangulating by minimizing reprojection error*

Start with a point $\mathbf{c}$ viewed in a calibrated camera and a corresponding point $\mathbf{c}'$ in a second calibrated camera. The rotation $\mathcal{R}$ and translation $\mathbf{t}$ from the first to the second camera are known. The first camera's intrinsic matrix is $\mathcal{K}_1$, etc. Write

$$\mathbf{m} = \mathcal{K}_1^{-1}\mathbf{c} \text{ and } \mathbf{m}' = \mathcal{K}_2^{-1}\mathbf{c}'$$

for the measurements in a standard camera. Now compute the reprojection error $E_r(\mathbf{X})$ for a variable 3D point $\mathbf{X}$ and minimize

$$E_r(\mathbf{X})$$

as a function of $\mathbf{X}$. Use a quasi-newton method for minimization.

33.2.3 The Essential Matrix

In the second camera's coordinate system, $\mathbf{v}_2' = \mathcal{R}^T\mathbf{v}_2 = \mathcal{R}^T(\mathbf{X} - \mathbf{t})$. The three vectors $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{t}$ must be coplanar. This means that

$$[\mathbf{t} \times \mathbf{v}_1]^T \mathbf{v}_2 = 0$$

A convenient trick from linear algebra helps here. For a vector $\mathbf{a} = [a_1, a_2, a_3]^T$, write

$$[\mathbf{a}]_X = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix}$$

and notice $\mathbf{a} \times \mathbf{b} = [\mathbf{a}]_X\mathbf{b}$. This means that

$$\begin{aligned} [[\mathbf{t}]_X\mathbf{v}_1]^T \mathbf{v}_2 &= 0 \\ &= \mathbf{v}_1^T[\mathbf{t}]_X{}^T\mathbf{v}_2 \\ &= \mathbf{v}_1^T[\mathbf{t}]_X{}^T\mathcal{R}\mathbf{v}_2' \end{aligned}$$

Use homogenous coordinates to express the point measured in camera 1 as $\mathbf{x}_1$, and notice there must be some constant $s \neq 0$ such that $\mathbf{x}_1 = s\mathbf{v}_1$. Similarly, use homogenous coordinates to express the point measured in camera 2 as $\mathbf{x}_2'$. Again, there must be some constant $t \neq 0$ such that $\mathbf{x}_2' = t\mathbf{v}_2'$. Because the equation above is homogenous,

$$\begin{aligned} \mathbf{x}_1^T[\mathbf{t}]_X{}^T\mathcal{R}\mathbf{x}_2' &= 0 \\ \mathbf{x}_1^T\mathcal{E}\mathbf{x}_2' &= 0 \end{aligned}$$

where $\mathcal{E}$ is known as the *essential matrix*.

33.2.4 Properties of the Essential Matrix

The essential matrix can be obtained *up to scale* from the fundamental matrix if you know the intrinsic calibration of the two cameras. The calibration matrices are written $\mathcal{K}_1$ and $\mathcal{K}_2$, and must have full rank. Then you can estimate the essential matrix as

$$\mathcal{K}_1^{-T} \mathcal{F} \mathcal{K}_2^{-1}.$$

Since the fundamental matrix is meaningful only up to scale, this estimate is only up to scale as well. This means that, if you apply the constructions above for the fundamental matrix to the essential matrix, they yield epipolar lines or epipoles *in the world coordinate system of the relevant camera.* The essential matrix must have rank 2 because the fundamental matrix does.

Recall that SVD($\mathcal{M}$) produces three matrices $\mathcal{U}_{\mathcal{M}}$, $\Sigma_{\mathcal{M}}$ and $\mathcal{V}_{\mathcal{M}}$, such that $\mathcal{M} = \mathcal{U}_{\mathcal{M}} \Sigma_{\mathcal{M}} \mathcal{V}_{\mathcal{M}}^T$ where $\mathcal{U}_{\mathcal{M}}$ and $\mathcal{V}_{\mathcal{M}}$ are orthonormal and $\Sigma_{\mathcal{M}}$ is diagonal. The terms on the diagonal of Σ are often referred to as *the singular values* of $\mathcal{M}$.

Recall

$$[\mathbf{t}]_X \mathbf{u} = \mathbf{t} \times \mathbf{u}.$$

This means

$$[\mathbf{t}]_X \mathbf{t} = \mathbf{0} \text{ and } \mathbf{t}^T [\mathbf{t}]_X = [\mathbf{t}]_X{}^T \mathbf{t} = -[\mathbf{t}]_X \mathbf{t} = \mathbf{0}$$

so $[\mathbf{t}]_X$ must have one singular value that is zero. The other two singular values are both $\| \mathbf{t} \|$, because

$$\| \mathbf{t} \times \mathbf{u} \| \leq \| \mathbf{t} \| \| \mathbf{u} \|$$

(with equality when $\mathbf{t}^T \mathbf{u} = 0$) and because there is a two-dimensional space of $\mathbf{u}$ that achieve equality. Because the essential matrix is known only up to scale, it is enough to require its singular values are 1, 1, and 0. Assume you are presented with an estimate of the essential matrix up to scale. This estimate can be corrected to be more like an essential matrix as below.

Procedure: 33.5 *Correcting an estimate of an essential matrix*

Given $\mathcal{M}$ – an estimate of an essential matrix – compute the SVD to obtain $\mathcal{U}$, Σ and $\mathcal{V}$. Compute $\Sigma_e = \text{diag}(1, 1, 0)$ and

$$\hat{\mathcal{E}} = \mathcal{U} \Sigma_e \mathcal{V}^T$$

which is the closest matrix to $\mathcal{M}$ that has appropriate singular values.

33.3 VISUAL ODOMETRY: EXPLOITING AN ESSENTIAL MATRIX

Remarkably, an essential matrix reveals the transformation between the two cameras up to scale. Recovering this information involves slightly complicated geometry. The result – given two images from cameras with known calibration, you can determine the relations between the cameras – is hugely useful.

33.3.1 Recovering Translation up to Scale

Assume you have two images obtained from two cameras whose calibration matrices you know. Further, you have the fundamental matrix up to scale, using the procedures of Section **??**. You can recover the essential matrix up to scale without difficulty, then correct it as in Procedure 33.5. From this, you can recover information about how the cameras moved from the essential matrix.

Recall for $\mathbf{t}$ the translation between cameras and $\mathcal{R}$ the rotation, the essential matrix $\hat{\mathcal{E}}$ has the property

$$s\hat{\mathcal{E}} = [\mathbf{t}]_X \mathcal{R}$$

(for any $s \neq 0$). Further $\mathbf{t}^T [\mathbf{t}]_X = \mathbf{0}$, so you can immediately recover the translation $\mathbf{t}$ *up to scale* by finding the unit vector $\mathbf{u}$ such that

$$\mathbf{u}^T \hat{\mathcal{E}} = \mathbf{0}^T.$$

These vector is occasionally referred to as the *left null vector* of the essential matrix. There are two unit left null vectors, which I write $\mathbf{u}_+$ and $\mathbf{u}_- = -\mathbf{u}_+$. Each is an estimate of the translation, so write $\hat{\mathbf{t}}_+$ and $\hat{\mathbf{t}}_-$ for the estimates. There are two distinct estimates of the translation.

33.3.2 Recovering the Rotation

The *right null vector* is the vector $\mathbf{v}$ such that

$$\hat{\mathcal{E}}\mathbf{v} = \mathbf{0}.$$

Again, there are two unit right null vectors, which I write $\mathbf{v}_+$ and $\mathbf{v}_- = -\mathbf{v}_+$. Each is an estimate of $\mathcal{R}^T \mathbf{t}$, which I write $\mathbf{s}$; for the two estimates, write $\hat{\mathbf{s}}_+$ and $\hat{\mathbf{s}}_-$. This leads to an important ambiguity in the rotation estimate. Assume, for the moment, you choose the translation estimate $\hat{\mathbf{t}}_+$. Then there are at least *two distinct* rotation estimates available. One estimate is obtained from $\hat{\mathbf{s}}_+ = \mathcal{R}_a \hat{\mathbf{t}}_+$ and the other from $\hat{\mathbf{s}}_- = \mathcal{R}_b \hat{\mathbf{t}}_+$.

Obtaining these rotation estimates takes some more work. Because $\hat{\mathcal{E}}$ has the right singular values and $\hat{\mathbf{t}}$ is the left null vector,

$$\hat{\mathcal{E}} = \left[\hat{\mathbf{t}}\right]_X \mathcal{R}$$

for an unknown $\mathcal{R}$. Now write

$$\hat{\mathcal{B}} = \left[\hat{\mathbf{t}}\right]_X^T \hat{\mathcal{E}} = \left[\hat{\mathbf{t}}\right]_X^T \left[\hat{\mathbf{t}}\right]_X \mathcal{R}$$

for a *known matrix* (you can multiply the estimates). SVD this matrix to get $\mathcal{U}_B$, Σ and $\mathcal{V}_B$, and SVD $[\mathbf{t}]_X$ to get $\mathcal{U}_T$, Σ_T and $\mathcal{V}_T$. Now

$$\begin{array}{rcl} \hat{\mathcal{B}} & = & \mathcal{U}_B \Sigma \mathcal{V}_B^T \\ & = & \mathcal{V}_T \Sigma_T^2 \mathcal{V}_T^{\mathcal{R}} \end{array}$$

and $\Sigma_T^2 = \Sigma_T = \Sigma$ which implies that you can estimate $\mathcal{R} = \mathcal{U}_B \mathcal{V}_B^T$. But $\Sigma = \mathrm{diag}(1, 1, 0)$ and this zero singular value creates some important ambiguities. First, notice that the last column of $\mathcal{U}_B$ must be parallel to $\hat{\mathbf{t}}$ to ensure that $\hat{\mathcal{B}}$ has the correct left null vector. Similarly, the last column of $\mathcal{V}_B$ must be parallel to $\hat{\mathbf{s}}$ to ensure that $\hat{\mathcal{B}}$ has the correct right null vector. There are now four possible options for the pairs $(\mathcal{U}_B, \mathcal{V}_B)$, obtained by choosing one of $\hat{\mathbf{t}}_+$ or $\hat{\mathbf{t}}_-$ and one of $\hat{\mathbf{s}}_+$ or $\hat{\mathbf{s}}_-$. Write $++, +-, -+, --$ for these options. Write $\mathcal{U}_{B,+} = \left[\alpha_1, \alpha_2, \hat{\mathbf{t}}_+\right]$ and $\mathcal{V}_{B,+} = [\beta_1, \beta_2, \hat{\mathbf{s}}_+]$, etc. Notice that $\mathcal{U}_{B,+}\mathcal{V}_{B,+}^T = \mathcal{U}_{B,-}\mathcal{V}_{B,-}^T$ and $\mathcal{U}_{B,-}\mathcal{V}_{B,+}^T = \mathcal{U}_{B,+}\mathcal{V}_{B,-}^T$, so there are two rotation options. Construct the two matrices

$$\mathcal{W}_+ = \alpha_1 \beta_1^T + \alpha_2 \beta_2^T - \hat{\mathbf{t}}\hat{\mathbf{s}}^T$$

and

$$\mathcal{W}_- = -\alpha_1 \beta_1^T - \alpha_2 \beta_2^T - \hat{\mathbf{t}}\hat{\mathbf{s}}^T.$$

and write

$$\hat{\mathcal{R}}_+ = \mathcal{W}_+(\det(\mathcal{W}_+))$$

$$\hat{\mathcal{R}}_- = \mathcal{W}_-(\det(\mathcal{W}_-)).$$

These are (a) true rotations, because their determinants are positive and (b) estimates of the rotation. The essential matrix yields four possible camera configurations:

$$(\hat{\mathbf{t}}_+, \hat{\mathcal{R}}_+), (\hat{\mathbf{t}}_-, \hat{\mathcal{R}}_+), (\hat{\mathbf{t}}_+, \hat{\mathcal{R}}_-), (\hat{\mathbf{t}}_-, \hat{\mathcal{R}}_-).$$

Procedure: 33.6 *Estimating camera rotation and translation from an essential matrix*

Given $\mathcal{E}$, an essential matrix, construct $\hat{\mathbf{t}}$, the unit left null vector, which is an estimate of the translation up to scale. Write $\hat{\mathcal{B}} = \left[\hat{\mathbf{t}}\right]_X^T \hat{\mathcal{E}}$, and compute the SVD to obtain $\mathcal{U}_B$, Σ and $\mathcal{V}_B$. Write $\mathcal{U}_B = [\alpha_1, \alpha_2, \mathbf{t}]$ and $\mathcal{V}_B = [\beta_1, \beta_2, \mathbf{s}]$ (where $\mathbf{s}$ is the unit right null vector). Construct the two matrices

$$\mathcal{W}_+ = \alpha_1 \beta_1^T + \alpha_2 \beta_2^T - \hat{\mathbf{t}}\hat{\mathbf{s}}^T$$

and

$$\mathcal{W}_- = -\alpha_1 \beta_1^T - \alpha_2 \beta_2^T - \hat{\mathbf{t}}\hat{\mathbf{s}}^T.$$

and write

$$\hat{\mathcal{R}}_+ = \mathcal{W}_+(\det(\mathcal{W}_+))$$

$$\hat{\mathcal{R}}_- = \mathcal{W}_-(\det(\mathcal{W}_-)).$$

The essential matrix yields four possible camera configurations:

$$(\hat{\mathbf{t}}, \hat{\mathcal{R}}_+), (-\hat{\mathbf{t}}_-, \hat{\mathcal{R}}_+), (\hat{\mathbf{t}}, \hat{\mathcal{R}}_-), (-\hat{\mathbf{t}}_-, \hat{\mathcal{R}}_-).$$

I have put the process in a box, above. While it appears to yield four solutions, only one is consistent with real imaging geometry.

33.3.3 Disambiguating Reconstructions

The relations between $\hat{\mathcal{R}}_+$ and $\hat{\mathcal{R}}_-$ are revealing. Compute

$$\hat{\mathcal{R}}_- \hat{\mathcal{R}}_+^T = -\alpha_1 \alpha_1^T - \alpha_2 \alpha_2^T + \hat{\mathbf{t}}\hat{\mathbf{t}}^T$$

and this maps $\mathbf{t}$ to $\mathbf{t}$, α_1 to $-\alpha_1$ and α_2 to $-\alpha_2$ – it is a rotation by 180^0 around the axis $\hat{\mathbf{t}}$. This makes it possible to visualize the four reconstructions. Figure 33.6 shows the four ambiguous reconstructions of camera 2 around camera 1. I have visualized the result of triangulating one point in each of the four reconstructions. In this case, the point is at the center of camera 1 and at the center of camera 2. Each of the camera reconstructions *has the same essential matrix*, so that if a point in one reconstruction corresponds to a point in camera 1, so does that point in each reconstruction. In turn, this means you can triangulate the point in each reconstruction with the point in camera 1. This leads to four distinct points in space, shown in the figure. Notice that, in this figure, only one of the four triangulated points lies in front of both camera 1 and a reconstruction. This is the general case. You can choose the correct reconstruction by this property, which you can test by looking at the sign of the Z-coordinate *in each camera's frame* **exercises** .

> **Procedure: 33.7** *Disambiguating odometry solutions*
>
> Construct the four solutions of Procedure 33.6, yielding four distinct cameras. Pair each of these cameras with the unit camera (which is camera 1) and for each of these four pairs, triangulate a set of points. Ideally, in one pair, the points will all be in front of both cameras. In practice, error in localizing the points might lead some to be behind one camera, so choose the reconstruction where the largest fraction of reconstructions is in front of both cameras.

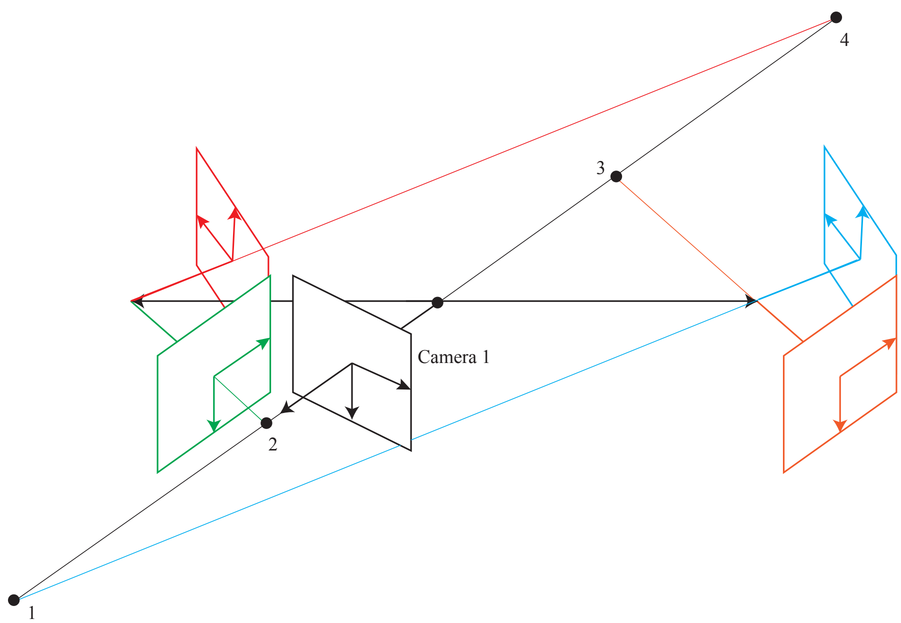

FIGURE 33.6: *There are four distinct possible reconstructions for the transformation from camera 1 to camera 2 implied by a known essential matrix. In this figure, camera 1 is at the canonical location, and the red, green, blue and orange cameras are the four reconstructions. Notice how red and green are related by a* 180^0 *rotation about the translation vector, as are blue and orange. The red-green pair are associated with one sign for the translation estimate (the black arrow), and the blue-orange pair are associated with the other sign. Each reconstruction has a triangulation of a point associated with it. Here I have used the point at each camera center* **exercises** *. I have shown the triangulated points; 1 corresponds to the blue camera; 3 to the orange camera; 2 to the green camera and 4 to the red camera. Notice how only one triangulation – in this case, 2 – produces a point that lies in front of both camera 1 and the reconstruction. In turn, the green camera is the correct reconstruction.*