Chapter 5

Reconstruction from Two Calibrated Views

We see because we move; we move because we see.
– James J. Gibson, *The Perception of the Visual World*

In this chapter we begin unveiling the basic geometry that relates images of points to their 3-D position. We start with the simplest case of two calibrated cameras, and describe an algorithm, first proposed by the British psychologist H.C. Longuet-Higgins in 1981, to reconstruct the relative pose (i.e. position and orientation) of the cameras as well as the locations of the points in space from their projection onto the two images.

It has been long known in photogrammetry that the coordinates of the projection of a point and the two camera optical centers form a triangle (Figure 5.1), a fact that can be written as an algebraic constraint involving the camera poses and image coordinates but *not* the 3-D position of the points. Given enough points, therefore, this constraint can be used to solve for the camera poses. Once those are known, the 3-D position of the points can be obtained easily by triangulation. The interesting feature of the constraint is that although it is nonlinear in the unknown camera poses, it can be solved by two linear steps in closed form. Therefore, in the absence of any noise or uncertainty, given two images taken from calibrated cameras, one can in principle recover camera pose and position of the points in space with a few steps of simple linear algebra.

While we have not yet indicated how to calibrate the cameras (which we will do in Chapter 6), this chapter serves to introduce the basic building blocks of the geometry of two views, known as "epipolar geometry." The simple algorithms to be introduced in this chapter, although merely conceptual,[1] allow us to introduce the basic ideas that will be revisited in later chapters of the book to derive more powerful algorithms that can deal with uncertainty in the measurements as well as with uncalibrated cameras.

5.1 Epipolar geometry

Consider two images of the same scene taken from two distinct vantage points. If we assume that the camera is *calibrated*, as described in Chapter 3 (the calibration matrix K is the identity), the homogeneous image coordinates $\boldsymbol{x}$ and the spatial coordinates $\boldsymbol{X}$ of a point p, with respect to the camera frame, are related by[2]

$$\lambda \boldsymbol{x} = \Pi_0 \boldsymbol{X}, \tag{5.1}$$

where $\Pi_0 = [I, 0]$. That is, the image $\boldsymbol{x}$ differs from the actual 3-D coordinates of the point by an unknown (depth) scale $\lambda \in \mathbb{R}_+$. For simplicity, we will assume that the scene is *static* (that is, there are no moving objects) and that the position of corresponding feature points across images is available, for instance from one of the algorithms described in Chapter 4. If we call $\boldsymbol{x}_1, \boldsymbol{x}_2$ the corresponding points in two views, they will then be related by a precise geometric relationship that we describe in this section.

5.1.1 The epipolar constraint and the essential matrix

Following Chapter 3, an orthonormal reference frame is associated with each camera, with its origin o at the optical center and the z-axis aligned with the optical axis. The relationship between the 3-D coordinates of a point in the inertial "world" coordinate frame and the camera frame can be expressed by a rigid-body transformation. Without loss of generality, we can assume the world frame to be one of the cameras, while the other is positioned and oriented according to a Euclidean transformation $g = (R, T) \in SE(3)$. If we call the 3-D coordinates of a point p relative to the two camera frames $\boldsymbol{X}_1 \in \mathbb{R}^3$ and $\boldsymbol{X}_2 \in \mathbb{R}^3$, they are related by a rigid-body transformation in the following way:

$$\boldsymbol{X}_2 = R\boldsymbol{X}_1 + T.$$

Now let $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^3$ be the homogeneous coordinates of the projection of *the same* point p in the two image planes. Since $\boldsymbol{X}_i = \lambda_i \boldsymbol{x}_i, i = 1, 2$, this equation

[1] They are not suitable for real images, which are typically corrupted by noise. In Section 5.2.3 of this chapter, we show how to modify them so as to minimize the effect of noise and obtain an optimal solution.

[2] We remind the reader that we do not distinguish between ordinary and homogeneous coordinates; in the former case $\boldsymbol{x} \in \mathbb{R}^2$, whereas in the latter $\boldsymbol{x} \in \mathbb{R}^3$ with the last component being 1. Similarly, $\boldsymbol{X} \in \mathbb{R}^3$ or $\boldsymbol{X} \in \mathbb{R}^4$ depends on whether ordinary or homogeneous coordinates are used.

can be written in terms of the image coordinates $\boldsymbol{x}_i$ and the depths λ_i as

$$\lambda_2 \boldsymbol{x}_2 = R\lambda_1 \boldsymbol{x}_1 + T.$$

In order to eliminate the depths λ_i in the preceding equation, premultiply both sides by $\widehat{T}$ to obtain

$$\lambda_2 \widehat{T} \boldsymbol{x}_2 = \widehat{T} R \lambda_1 \boldsymbol{x}_1.$$

Since the vector $\widehat{T}\boldsymbol{x}_2 = T \times \boldsymbol{x}_2$ is perpendicular to the vector $\boldsymbol{x}_2$, the inner product $\langle \boldsymbol{x}_2, \widehat{T}\boldsymbol{x}_2 \rangle = \boldsymbol{x}_2{}^T \widehat{T} \boldsymbol{x}_2$ is zero. Premultiplying the previous equation by $\boldsymbol{x}_2^T$ yields that the quantity $\boldsymbol{x}_2^T \widehat{T} R \lambda_1 \boldsymbol{x}_1$ is zero. Since $\lambda_1 > 0$, we have proven the following result:

Theorem 5.1 (Epipolar constraint). *Consider two images $\boldsymbol{x}_1, \boldsymbol{x}_2$ of the same point p from two camera positions with relative pose (R, T), where $R \in SO(3)$ is the relative orientation and $T \in \mathbb{R}^3$ is the relative position. Then $\boldsymbol{x}_1, \boldsymbol{x}_2$ satisfy*

$$\langle \boldsymbol{x}_2, T \times R\boldsymbol{x}_1 \rangle = 0, \quad \text{or} \quad \boxed{\boldsymbol{x}_2^T \widehat{T} R \boldsymbol{x}_1 = 0.} \tag{5.2}$$

The matrix

$$E \doteq \widehat{T} R \quad \in \mathbb{R}^{3\times 3}$$

in the epipolar constraint equation (5.2) is called the *essential matrix*. It encodes the relative pose between the two cameras. The epipolar constraint (5.2) is also called the *essential constraint*. Since the epipolar constraint is bilinear in each of its arguments $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, it is also called the *bilinear constraint*. We will revisit this bilinear nature in later chapters.

In addition to the preceding algebraic derivation, this constraint follows immediately from geometric considerations, as illustrated in Figure 5.1. The vector connecting the first camera center o_1 and the point p, the vector connecting o_2

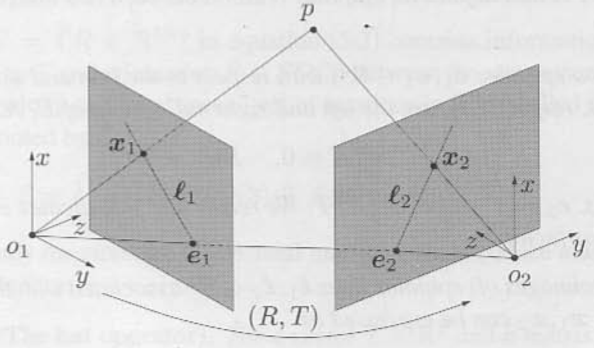

Figure 5.1. Two projections $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^3$ of a 3-D point p from two vantage points. The Euclidean transformation between the two cameras is given by $(R, T) \in SE(3)$. The intersections of the line (o_1, o_2) with each image plane are called *epipoles* and denoted by $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$. The lines $\boldsymbol{\ell}_1, \boldsymbol{\ell}_2$ are called *epipolar lines*, which are the intersection of the plane (o_1, o_2, p) with the two image planes.

and p, and the vector connecting the two optical centers o_1 and o_2 clearly form a triangle. Therefore, the three vectors lie on the same plane. Their triple product,[3] which measures the volume of the parallelepiped determined by the three vectors, is therefore zero. This is true for the coordinates of the points $\boldsymbol{X}_i,\ i = 1, 2$, as well as for the homogeneous coordinates of their projection $\boldsymbol{x}_i,\ i = 1, 2$, since $\boldsymbol{X}_i$ and $\boldsymbol{x}_i$ (as vectors) differ only be a scalar factor. The constraint (5.2) is just the triple product written in the second camera frame; $R\boldsymbol{x}_1$ is simply the direction of the vector $\overrightarrow{o_1 p}$, and T is the vector $\overrightarrow{o_2 o_1}$ with respect to the second camera frame. The translation T between the two camera centers o_1 and o_2 is also called the *baseline*.

Associated with this picture, we define the following set of geometric entities, which will facilitate our future study:

Definition 5.2 (Epipolar geometric entities).

1. *The plane (o_1, o_2, p) determined by the two centers of projection o_1, o_2 and the point p is called an* epipolar plane *associated with the camera configuration and point p. There is one epipolar plane for each point p.*
2. *The projection $\mathbf{e}_1(\mathbf{e}_2)$ of one camera center onto the image plane of the other camera frame is called an* epipole. *Note that the projection may occur outside the physical boundary of the imaging sensor.*
3. *The intersection of the epipolar plane of p with one image plane is a line $\boldsymbol{\ell}_1(\boldsymbol{\ell}_2)$, which is called the* epipolar line *of p. We usually use the normal vector $\boldsymbol{\ell}_1(\boldsymbol{\ell}_2)$ to the epipolar plane to denote this line.*[4]

From the definitions, we immediately have the following relations among epipoles, epipolar lines, and image points:

Proposition 5.3 (Properties of epipoles and epipolar lines). *Given an essential matrix $E = \widehat{T}R$ that defines an epipolar relation between two images $\boldsymbol{x}_1, \boldsymbol{x}_2$, we have:*

1. *The two epipoles $\mathbf{e}_1, \mathbf{e}_2 \in \mathbb{R}^3$, with respect to the first and second camera frames, respectively, are the left and right null spaces of E, respectively:*
$$\mathbf{e}_2^T E = 0, \quad E\mathbf{e}_1 = 0. \tag{5.3}$$
That is, $\mathbf{e}_2 \sim T$ and $\mathbf{e}_1 \sim R^T T$. We recall that $\sim$ indicates equality up to a scalar factor.
2. *The (coimages of) epipolar lines $\boldsymbol{\ell}_1, \boldsymbol{\ell}_2 \in \mathbb{R}^3$ associated with the two image points $\boldsymbol{x}_1, \boldsymbol{x}_2$ can be expressed as*
$$\boldsymbol{\ell}_2 \sim E\boldsymbol{x}_1, \quad \boldsymbol{\ell}_1 \sim E^T\boldsymbol{x}_2 \quad \in \mathbb{R}^3, \tag{5.4}$$

[3] As we have seen in Chapter 2, the triple product of three vectors is the inner product of one with the cross product of the other two.

[4] Hence the vector $\boldsymbol{\ell}_1(\boldsymbol{\ell}_2)$ is in fact the coimage of the epipolar line.

where $\boldsymbol{\ell}_1, \boldsymbol{\ell}_2$ are in fact the normal vectors to the epipolar plane expressed with respect to the two camera frames, respectively.

3. *In each image, both the image point and the epipole lie on the epipolar line*

$$\boldsymbol{\ell}_i^T e_i = 0, \quad \boldsymbol{\ell}_i^T \boldsymbol{x}_i = 0, \quad i = 1, 2. \tag{5.5}$$

The proof is simple, and we leave it to the reader as an exercise. Figure 5.2 illustrates the relationships among 3-D points, images, epipolar lines, and epipoles.

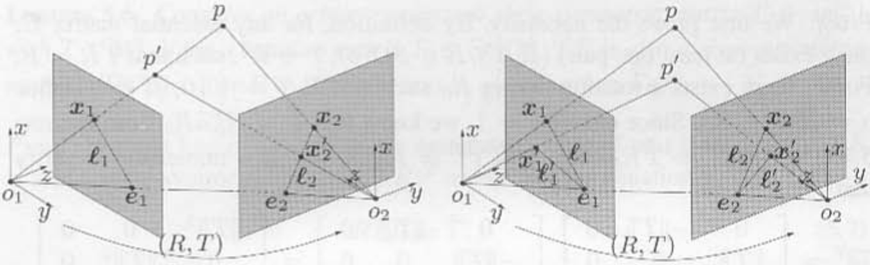

Figure 5.2. Left: the essential matrix E associated with the epipolar constraint maps an image point $\boldsymbol{x}_1$ in the first image to an epipolar line $\boldsymbol{\ell}_2 = E\boldsymbol{x}_1$ in the second image; the precise location of its corresponding image ($\boldsymbol{x}_2$ or $\boldsymbol{x}_2'$) depends on where the 3-D point (p or p') lies on the ray $(o_1, \boldsymbol{x}_1)$; Right: When (o_1, o_2, p) and (o_1, o_2, p') are two different planes, they intersect at the two image planes at two pairs of epipolar lines $(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2)$ and $(\boldsymbol{\ell}_1', \boldsymbol{\ell}_2')$, respectively, and these epipolar lines always pass through the pair of epipoles (e_1, e_2).

5.1.2 *Elementary properties of the essential matrix*

The matrix $E = \widehat{T}R \in \mathbb{R}^{3\times3}$ in equation (5.2) contains information about the relative position T and orientation $R \in SO(3)$ between the two cameras. Matrices of this form belong to a very special set of matrices in $\mathbb{R}^{3\times3}$ called the *essential space* and denoted by $\mathcal{E}$:

$$\mathcal{E} \doteq \left\{ \widehat{T}R \,\middle|\, R \in SO(3), T \in \mathbb{R}^3 \right\} \subset \mathbb{R}^{3\times3}.$$

Before we study the structure of essential matrices, we introduce a useful lemma from linear algebra.

Lemma 5.4 (The hat operator). *For a vector $T \in \mathbb{R}^3$ and a matrix $K \in \mathbb{R}^{3\times3}$, if $\det(K) = +1$ and $T' = KT$, then $\widehat{T} = K^T\widehat{T'}K$.*

Proof. Since both $K^T\widehat{(\cdot)}K$ and $\widehat{K^{-1}(\cdot)}$ are linear maps from $\mathbb{R}^3$ to $\mathbb{R}^{3\times3}$, one may directly verify that these two linear maps agree on the basis vectors $[1,0,0]^T$, $[0,1,0]^T$, and $[0,0,1]^T$ (using the fact that $\det(K) = 1$). □

The following theorem, due to [Huang and Faugeras, 1989], captures the algebraic structure of essential matrices in terms of their singular value decomposition (see Appendix A for a review on the SVD):

Theorem 5.5 (Characterization of the essential matrix). *A nonzero matrix $E \in \mathbb{R}^{3\times 3}$ is an essential matrix* if and only if *E has a singular value decomposition (SVD) $E = U\Sigma V^T$ with*

$$\Sigma = diag\{\sigma, \sigma, 0\}$$

for some $\sigma \in \mathbb{R}_+$ and $U, V \in SO(3)$.

Proof. We first prove the necessity. By definition, for any essential matrix E, there exists (at least one pair) $(R, T), R \in SO(3), T \in \mathbb{R}^3$, such that $\widehat{T}R = E$. For T, there exists a rotation matrix R_0 such that $R_0 T = [0, 0, \|T\|]^T$. Define $a = R_0 T \in \mathbb{R}^3$. Since $\det(R_0) = 1$, we know that $\widehat{T} = R_0^T \widehat{a} R_0$ from Lemma 5.4. Then $EE^T = \widehat{T}RR^T\widehat{T}^T = \widehat{T}\widehat{T}^T = R_0^T \widehat{a}\widehat{a}^T R_0$. It is immediate to verify that

$$\widehat{a}\widehat{a}^T = \begin{bmatrix} 0 & -\|T\| & 0 \\ \|T\| & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & \|T\| & 0 \\ -\|T\| & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \|T\|^2 & 0 & 0 \\ 0 & \|T\|^2 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

So, the singular values of the essential matrix $E = \widehat{T}R$ are $(\|T\|, \|T\|, 0)$. However, in the standard SVD of $E = U\Sigma V^T$, U and V are only orthonormal, and their determinants can be ± 1.[5] We still need to prove that $U, V \in SO(3)$ (i.e. they have determinant $+1$) to establish the theorem. We already have $E = \widehat{T}R = R_0^T \widehat{a} R_0 R$. Let $R_Z(\theta)$ be the matrix that represents a rotation around the Z-axis by an angle of θ radians; i.e. $R_Z(\theta) \doteq e^{\widehat{e}_3\theta}$ with $e_3 = [0, 0, 1]^T \in \mathbb{R}^3$. Then

$$R_Z\left(+\frac{\pi}{2}\right) = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Then $\widehat{a} = R_Z(+\frac{\pi}{2})R_Z^T(+\frac{\pi}{2})\widehat{a} = R_Z(+\frac{\pi}{2})\ \mathrm{diag}\{\|T\|, \|T\|, 0\}$. Therefore,

$$E = \widehat{T}R = R_0^T R_Z\left(+\frac{\pi}{2}\right)\ \mathrm{diag}\{\|T\|, \|T\|, 0\} R_0 R.$$

So, in the SVD of $E = U\Sigma V^T$, we may choose $U = R_0^T R_Z(+\frac{\pi}{2})$ and $V^T = R_0 R$. Since we have constructed both U and V as products of matrices in $SO(3)$, they are in $SO(3)$, too; that is, both U and V are rotation matrices.

We now prove sufficiency. If a given matrix $E \in \mathbb{R}^{3\times 3}$ has SVD $E = U\Sigma V^T$ with $U, V \in SO(3)$ and $\Sigma = \mathrm{diag}\{\sigma, \sigma, 0\}$, define $(R_1, T_1) \in SE(3)$ and $(R_2, T_2) \in SE(3)$ to be

$$\left\{ \begin{array}{lcl} (\widehat{T}_1, R_1) & = & (UR_Z(+\frac{\pi}{2})\Sigma U^T, UR_Z^T(+\frac{\pi}{2})V^T), \\ (\widehat{T}_2, R_2) & = & (UR_Z(-\frac{\pi}{2})\Sigma U^T, UR_Z^T(-\frac{\pi}{2})V^T). \end{array} \right. \tag{5.6}$$

[5] Interested readers can verify this using the Matlab routine: SVD.

It is now easy to verify that $\widehat{T}_1 R_1 = \widehat{T}_2 R_2 = E$. Thus, E is an essential matrix. □

Given a rotation matrix $R \in SO(3)$ and a translation vector $T \in \mathbb{R}^3$, it is immediate to construct an essential matrix $E = \widehat{T}R$. The inverse problem, that is how to retrieve T and R from a given essential matrix E, is less obvious. In the sufficiency proof for the above theorem, we have used the SVD to construct two solutions for (R, T). Are these the only solutions? Before we can answer this question in the upcoming Theorem 5.7, we need the following lemma.

Lemma 5.6. *Consider an arbitrary nonzero skew-symmetric matrix $\widehat{T} \in so(3)$ with $T \in \mathbb{R}^3$. If for a rotation matrix $R \in SO(3)$, $\widehat{T}R$ is also a skew-symmetric matrix, then $R = I$ or $R = e^{\widehat{u}\pi}$, where $u = \frac{T}{\|T\|}$. Further, $\widehat{T}e^{\widehat{u}\pi} = -\widehat{T}$.*

Proof. Without loss of generality, we assume that T is of unit length. Since $\widehat{T}R$ is also a skew-symmetric matrix, $(\widehat{T}R)^T = -\widehat{T}R$. This equation gives

$$R\widehat{T}R = \widehat{T}. \tag{5.7}$$

Since R is a rotation matrix, there exist $\omega \in \mathbb{R}^3$, $\|\omega\| = 1$, and $\theta \in \mathbb{R}$ such that $R = e^{\widehat{\omega}\theta}$. If $\theta = 0$ the lemma is proved. Hence consider the case $\theta \neq 0$. Then, (5.7) is rewritten as $e^{\widehat{\omega}\theta}\widehat{T}e^{\widehat{\omega}\theta} = \widehat{T}$. Applying this equation to ω, we get $e^{\widehat{\omega}\theta}\widehat{T}e^{\widehat{\omega}\theta}\omega = \widehat{T}\omega$. Since $e^{\widehat{\omega}\theta}\omega = \omega$, we obtain $e^{\widehat{\omega}\theta}\widehat{T}\omega = \widehat{T}\omega$. Since ω is the only eigenvector associated with the eigenvalue 1 of the matrix $e^{\widehat{\omega}\theta}$, and $\widehat{T}\omega$ is orthogonal to ω, $\widehat{T}\omega$ has to be zero. Thus, ω is equal to either $\frac{T}{\|T\|}$ or $-\frac{T}{\|T\|}$; i.e. $\omega = \pm u$. Then R has the form $e^{\widehat{\omega}\theta}$, which commutes with $\widehat{T}$. Thus from (5.7), we get

$$e^{2\widehat{\omega}\theta}\widehat{T} = \widehat{T}. \tag{5.8}$$

According to *Rodrigues' formula* (2.16) from Chapter 2, we have

$$e^{2\widehat{\omega}\theta} = I + \widehat{\omega}\sin(2\theta) + \widehat{\omega}^2(1 - \cos(2\theta)),$$

and (5.8) yields

$$\widehat{\omega}^2\sin(2\theta) + \widehat{\omega}^3(1 - \cos(2\theta)) = 0.$$

Since $\widehat{\omega}^2$ and $\widehat{\omega}^3$ are linearly independent (we leave this as an exercise to the reader), we have $\sin(2\theta) = 1 - \cos(2\theta) = 0$. That is, θ is equal to $2k\pi$ or $2k\pi + \pi$, $k \in \mathbb{Z}$. Therefore, R is equal to I or $e^{\widehat{\omega}\pi}$. Now if $\omega = u = \frac{T}{\|T\|}$, then it is direct from the geometric meaning of the rotation $e^{\widehat{\omega}\pi}$ that $e^{\widehat{\omega}\pi}\widehat{T} = -\widehat{T}$. On the other hand, if $\omega = -u = -\frac{T}{\|T\|}$, then it follows that $e^{\widehat{\omega}}\widehat{T} = -\widehat{T}$. Thus, in any case the conclusions of the lemma follow. □

The following theorem shows exactly how many rotation and translation pairs (R, T) one can extract from an essential matrix, and the solutions are given in closed form by equation (5.9).

Theorem 5.7 (Pose recovery from the essential matrix). *There exist exactly two relative poses (R, T) with $R \in SO(3)$ and $T \in \mathbb{R}^3$ corresponding to a nonzero essential matrix $E \in \mathcal{E}$.*

Proof. Assume that $(R_1, T_1) \in SE(3)$ and $(R_2, T_2) \in SE(3)$ are both solutions for the equation $\widehat{T}R = E$. Then we have $\widehat{T}_1 R_1 = \widehat{T}_2 R_2$. This yields $\widehat{T}_1 = \widehat{T}_2 R_2 R_1^T$. Since $\widehat{T}_1, \widehat{T}_2$ are both skew-symmetric matrices and $R_2 R_1^T$ is a rotation matrix, from the preceding lemma, we have that either $(R_2, T_2) = (R_1, T_1)$ or $(R_2, T_2) = (e^{\widehat{u}_1 \pi} R_1, -T_1)$ with $u_1 = T_1/\|T_1\|$. Therefore, given an essential matrix E there are exactly *two* pairs of (R, T) such that $\widehat{T}R = E$. Further, if E has the SVD: $E = U\Sigma V^T$ with $U, V \in SO(3)$, the following formulae give the two distinct solutions (recall that $R_Z(\theta) \doteq e^{\widehat{e}_3 \theta}$ with $e_3 = [0,0,1]^T \in \mathbb{R}^3$)

$$\boxed{\begin{array}{rcl} (\widehat{T}_1, R_1) & = & \left(UR_Z(+\frac{\pi}{2})\Sigma U^T, UR_Z^T(+\frac{\pi}{2})V^T\right), \\ (\widehat{T}_2, R_2) & = & \left(UR_Z(-\frac{\pi}{2})\Sigma U^T, UR_Z^T(-\frac{\pi}{2})V^T\right). \end{array}} \tag{5.9}$$

□

Example 5.8 (Two solutions to an essential matrix). It is immediate to verify that $\widehat{e}_3 R_Z\left(+\frac{\pi}{2}\right) = \widehat{-e_3} R_Z\left(-\frac{\pi}{2}\right)$, since

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

These two solutions together are usually referred to as a "twisted pair," due to the manner in which the two solutions are related geometrically, as illustrated in Figure 5.3. A physically correct solution can be chosen by enforcing that the reconstructed points be visible, i.e. they have positive depth. We discuss this issue further in Exercise 5.11. ■

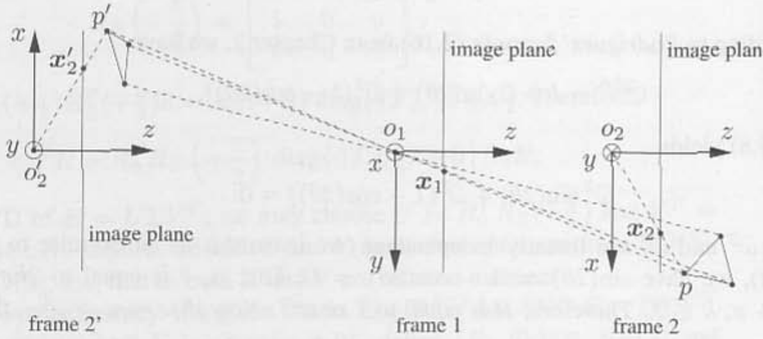

Figure 5.3. Two pairs of camera frames, i.e. $(1,2)$ and $(1,2')$, generate the same essential matrix. The frame 2 and frame $2'$ differ by a translation and a 180° rotation (a twist) around the z-axis, and the two pose pairs give rise to the same image coordinates. For the same set of image pairs x_1 and $x_2 = x_2'$, the recovered structures p and p' might be different. Notice that with respect to the camera frame 1, the point p' has a negative depth.

5.2 Basic reconstruction algorithms

In the previous section, we have seen that images of corresponding points are related by the epipolar constraint, which involves the unknown relative pose between the cameras. Therefore, given a number of corresponding points, we could use the epipolar constraints to try to recover camera pose. In this section, we show a simple closed-form solution to this problem. It consists of two steps: First a matrix E is recovered from a number of epipolar constraints; then relative translation and orientation are extracted from E. However, since the matrix E recovered using correspondence data in the epipolar constraint may not be an essential matrix, it needs to be projected into the space of essential matrices prior to extraction of the relative pose of the cameras using equation (5.9).

Although the linear algorithm that we propose here is suboptimal when the measurements are corrupted by noise, it is important for illustrating the geometric structure of the space of essential matrices. We leave the more practical issues with noise and optimality to Section 5.2.3.

5.2.1 *The eight-point linear algorithm*

Let $E = \widehat{T}R$ be the essential matrix associated with the epipolar constraint (5.2). The entries of this 3×3 matrix are denoted by

$$E = \begin{bmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{bmatrix} \quad \in \mathbb{R}^{3\times 3} \tag{5.10}$$

and stacked into a vector $E^s \in \mathbb{R}^9$, which is typically referred to as the *stacked version* of the matrix E (Appendix A.1.3):

$$E^s \doteq [e_{11}, e_{21}, e_{31}, e_{12}, e_{22}, e_{32}, e_{13}, e_{23}, e_{33}]^T \quad \in \mathbb{R}^9.$$

The inverse operation from E^s to its matrix version is then called *unstacking*. We further denote the *Kronecker product* $\otimes$ (also see Appendix A.1.3) of two vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ by

$$\boldsymbol{a} \doteq \boldsymbol{x}_1 \otimes \boldsymbol{x}_2. \tag{5.11}$$

Or, more specifically, if $\boldsymbol{x}_1 = [x_1, y_1, z_1]^T \in \mathbb{R}^3$ and $\boldsymbol{x}_2 = [x_2, y_2, z_2]^T \in \mathbb{R}^3$, then

$$\boldsymbol{a} = [x_1x_2, x_1y_2, x_1z_2, y_1x_2, y_1y_2, y_1z_2, z_1x_2, z_1y_2, z_1z_2]^T \quad \in \mathbb{R}^9. \tag{5.12}$$

Since the epipolar constraint $\boldsymbol{x}_2^T E \boldsymbol{x}_1 = 0$ is linear in the entries of E, using the above notation we can rewrite it as the inner product of $\boldsymbol{a}$ and E^s:

$$\boxed{\boldsymbol{a}^T E^s = 0.}$$

This is just another way of writing equation (5.2) that emphasizes the linear dependence of the epipolar constraint on the elements of the essential matrix. Now, given a set of corresponding image points $(\boldsymbol{x}_1^j, \boldsymbol{x}_2^j)$, $j = 1, 2, \ldots, n$, define a matrix $\chi \in \mathbb{R}^{n\times 9}$ associated with these measurements to be

$$\chi \doteq [\boldsymbol{a}^1, \boldsymbol{a}^2, \ldots, \boldsymbol{a}^n]^T, \tag{5.13}$$

where the jth row $\boldsymbol{a}^j$ is the Kronecker product of each pair $(\boldsymbol{x}_1^j, \boldsymbol{x}_2^j)$ using (5.12). In the absence of noise, the vector E^s satisfies

$$\chi E^s = 0. \tag{5.14}$$

This linear equation may now be solved for the vector E^s. For the solution to be unique (up to a scalar factor, ruling out the trivial solution $E^s = 0$), the rank of the matrix $\chi \in \mathbb{R}^{9\times n}$ needs to be exactly eight. This should be the case given $n \geq 8$ "ideal" corresponding points, as shown in Figure 5.4. In general, however, since correspondences may be prone to errors, there may be no solution to (5.14). In such a case, one can choose the E^s that minimizes the least-squares error function $\|\chi E^s\|^2$. This is achieved by choosing E^s to be the eigenvector of $\chi^T\chi$ that corresponds to its smallest eigenvalue, as we show in Appendix A. We would also like to draw attention to the case when the rank of χ is less then eight even for number of points greater than nine. In this instance there are multiple solutions to (5.14). This happens when the feature points are not in "general position," for example when they all lie on a plane. We will specifically deal with the planar case in the next section.

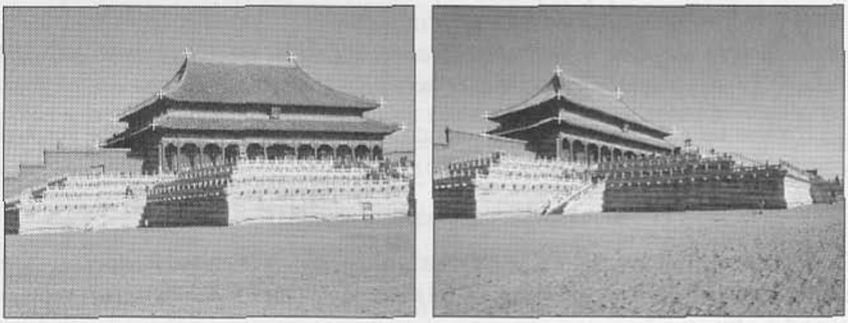

Figure 5.4. Eight pairs of corresponding image points in two views of the Tai-He palace in the Forbidden City, Beijing, China (photos courtesy of Jie Zhang).

However, even in the absence of noise, for a vector E^s to be the solution of our problem, it is not sufficient that it be in the null space of χ. In fact, it has to satisfy an additional constraint, that its matrix form E belong to the space of essential matrices. Enforcing this structure in the determination of the null space of χ is difficult. Therefore, as a first cut, we estimate the null space of χ, *ignoring the internal structure of essential matrix*, obtaining a matrix, say F, that probably does not belong to the essential space $\mathcal{E}$, and then "orthogonally" project the matrix thus obtained onto the essential space. This process is illustrated in Figure 5.5. The following theorem says precisely what this projection is.

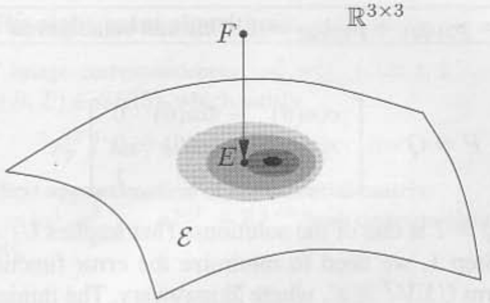

Figure 5.5. Among all points in the essential space $\mathcal{E} \subset \mathbb{R}^{3\times 3}$, E has the shortest Frobenius distance to F. However, the least-square error may not be the smallest for so-obtained E among all points in $\mathcal{E}$.

Theorem 5.9 (Projection onto the essential space). *Given a real matrix $F \in \mathbb{R}^{3\times 3}$ with SVD $F = U diag\{\lambda_1, \lambda_2, \lambda_3\} V^T$ with $U, V \in SO(3)$, $\lambda_1 \geq \lambda_2 \geq \lambda_3$, then the essential matrix $E \in \mathcal{E}$ that minimizes the error $\|E - F\|_f^2$ is given by $E = U diag\{\sigma, \sigma, 0\} V^T$ with $\sigma = (\lambda_1 + \lambda_2)/2$. The subscript "$f$" indicates the Frobenius norm of a matrix. This is the square norm of the sum of the squares of all the entries of the matrix (see Appendix A).*

Proof. For any fixed matrix $\Sigma = \text{diag}\{\sigma, \sigma, 0\}$, we define a subset $\mathcal{E}_\Sigma$ of the essential space $\mathcal{E}$ to be the set of all essential matrices with SVD of the form $U_1 \Sigma V_1^T$, $U_1, V_1 \in SO(3)$. To simplify the notation, define $\Sigma_\lambda = \text{diag}\{\lambda_1, \lambda_2, \lambda_3\}$. We now prove the theorem in two steps:

Step 1: We prove that for a fixed Σ, the essential matrix $E \in \mathcal{E}_\Sigma$ that minimizes the error $\|E - F\|_f^2$ has a solution $E = U\Sigma V^T$ (not necessarily unique). Since $E \in \mathcal{E}_\Sigma$ has the form $E = U_1 \Sigma V_1^T$, we get

$$\|E - F\|_f^2 = \|U_1 \Sigma V_1^T - U \Sigma_\lambda V^T\|_f^2 = \|\Sigma_\lambda - U^T U_1 \Sigma V_1^T V\|_f^2.$$

Define $P = U^T U_1, Q = V^T V_1 \in SO(3)$, which have the form

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix}, \quad Q = \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{bmatrix}. \tag{5.15}$$

Then

$$\begin{aligned} \|E - F\|_f^2 &= \|\Sigma_\lambda - U^T U_1 \Sigma V_1^T V\|_f^2 \\ &= \text{trace}(\Sigma_\lambda^2) - 2\text{trace}(P \Sigma Q^T \Sigma_\lambda) + \text{trace}(\Sigma^2). \end{aligned}$$

Expanding the second term, using $\Sigma = \text{diag}\{\sigma, \sigma, 0\}$ and the notation p_{ij}, q_{ij} for the entries of P, Q, we have

$$\text{trace}(P \Sigma Q^T \Sigma_\lambda) = \sigma\big(\lambda_1(p_{11}q_{11} + p_{12}q_{12}) + \lambda_2(p_{21}q_{21} + p_{22}q_{22})\big).$$

Since P, Q are rotation matrices, $p_{11}q_{11} + p_{12}q_{12} \leq 1$ and $p_{21}q_{21} + p_{22}q_{22} \leq 1$. Since Σ, Σ_λ are fixed and $\lambda_1, \lambda_2 \geq 0$, the error $\|E - F\|_f^2$ is minimized when

$p_{11}q_{11} + p_{12}q_{12} = p_{21}q_{21} + p_{22}q_{22} = 1$. This can be achieved when P, Q are of the general form

$$P = Q = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Obviously, $P = Q = I$ is one of the solutions. That implies $U_1 = U, V_1 = V$.

Step 2: From Step 1, we need to minimize the error function only over the matrices of the form $U\Sigma V^T \in \mathcal{E}$, where Σ may vary. The minimization problem is then converted to one of minimizing the error function

$$\|E - F\|_f^2 = (\lambda_1 - \sigma)^2 + (\lambda_2 - \sigma)^2 + (\lambda_3 - 0)^2.$$

Clearly, the σ that minimizes this error function is given by $\sigma = (\lambda_1 + \lambda_2)/2$. □

As we have already pointed out, the epipolar constraint allows us to recover the essential matrix only up to a scalar factor (since the epipolar constraint (5.2) is homogeneous in E, it is not modified by multiplying it by any nonzero constant). A typical choice to fix this ambiguity is to assume a unit translation, that is, $\|T\| = \|E\| = 1$. We call the resulting essential matrix *normalized*.

Remark 5.10. *The reader may have noticed that the above theorem relies on a special assumption that in the SVD of E both matrices U and V are rotation matrices in $SO(3)$. This is not always true when E is estimated from noisy data. In fact, standard SVD routines do* not *guarantee that the computed U and V have positive determinant. The problem can be easily resolved once one notices that the sign of the essential matrix E is also arbitrary (even after normalization). The above projection can operate either on $+E$ or $-E$. We leave it as an exercise to the reader that one of the (noisy) matrices $\pm E$ will always have an SVD that satisfies the conditions of Theorem 5.9.*

According to Theorem 5.7, each normalized essential matrix E gives two possible poses (R, T). So from $\pm E$, we can recover the pose up to four solutions. In fact, three of the solutions can be eliminated by imposing the positive depth constraint. We leave the details to the reader as an exercise (see Exercise 5.11).

The overall algorithm, which is due to [Longuet-Higgins, 1981], can then be summarized as Algorithm 5.1.

To account for the possible sign change $\pm E$, in the last step of the algorithm, the "$+$" and "$-$" signs in the equations for R and T should be arbitrarily combined so that all four solutions can be obtained.

Example 5.11 (A numerical example). Suppose that

$$R = \begin{bmatrix} \cos(\pi/4) & 0 & \sin(\pi/4) \\ 0 & 1 & 0 \\ -\sin(\pi/4) & 0 & \cos(\pi/4) \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \\ 0 & 1 & 0 \\ -\frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \end{bmatrix}, \quad T = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}.$$

Algorithm 5.1 (The eight-point algorithm).

For a given set of image correspondences (x_1^j, x_2^j), $j = 1, 2, \ldots, n$ $(n \geq 8)$, this algorithm recovers $(R, T) \in SE(3)$, which satisfy

$$x_2^{jT} \widehat{T} R x_1^j = 0, \quad j = 1, 2, \ldots, n.$$

1. **Compute a first approximation of the essential matrix**
   Construct $\chi = [a^1, a^2, \ldots, a^n]^T \in \mathbb{R}^{n \times 9}$ from correspondences x_1^j and x_2^j as in (5.12), namely,
   $$a^j = x_1^j \otimes x_2^j \quad \in \mathbb{R}^9.$$
   Find the vector $E^s \in \mathbb{R}^9$ of unit length such that $\|\chi E^s\|$ is minimized as follows: compute the SVD of $\chi = U_\chi \Sigma_\chi V_\chi^T$ and define E^s to be the ninth column of V_χ. Unstack the nine elements of E^s into a square 3×3 matrix E as in (5.10). Note that this matrix will in general *not* be in the essential space.

2. **Project onto the essential space**
   Compute the singular value decomposition of the matrix E recovered from data to be
   $$E = U \text{diag}\{\sigma_1, \sigma_2, \sigma_3\} V^T,$$
   where $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq 0$ and $U, V \in SO(3)$. In general, since E may not be an essential matrix, $\sigma_1 \neq \sigma_2$ and $\sigma_3 \neq 0$. But its projection onto the normalized essential space is $U \Sigma V^T$, where $\Sigma = \text{diag}\{1, 1, 0\}$.

3. **Recover the displacement from the essential matrix**
   We now need only U and V to extract R and T from the essential matrix as
   $$R = U R_Z^T\left(\pm\frac{\pi}{2}\right) V^T, \quad \widehat{T} = U R_Z\left(\pm\frac{\pi}{2}\right) \Sigma U^T.$$
   where $R_Z^T\left(\pm\frac{\pi}{2}\right) \doteq \begin{bmatrix} 0 & \pm 1 & 0 \\ \mp 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

Then the essential matrix is

$$E = \widehat{T} R = \begin{bmatrix} 0 & 0 & 0 \\ \sqrt{2} & 0 & -\sqrt{2} \\ 0 & 2 & 0 \end{bmatrix}.$$

Since $\|T\| = 2$, the E obtained here is not normalized. It is also easy to see this from its SVD,

$$E = U \Sigma V^T \doteq \begin{bmatrix} 0 & 0 & -1 \\ -1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -\frac{\sqrt{2}}{2} & 0 & -\frac{\sqrt{2}}{2} \\ 0 & 1 & 0 \\ \frac{\sqrt{2}}{2} & 0 & -\frac{\sqrt{2}}{2} \end{bmatrix}^T,$$

where the nonzero singular values are 2 instead of 1. Normalizing E is equivalent to replacing the above Σ by

$$\Sigma = \text{diag}\{1, 1, 0\}.$$

It is then easy to compute the four possible decompositions $(R, \widehat{T})$ for E:

$$1.\quad UR_Z^T\left(\frac{\pi}{2}\right)V^T = \begin{bmatrix} \frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \\ 0 & -1 & 0 \\ \frac{\sqrt{2}}{2} & 0 & -\frac{\sqrt{2}}{2} \end{bmatrix}, \quad UR_Z\left(\frac{\pi}{2}\right)\Sigma U^T = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix};$$

$$2.\quad UR_Z^T\left(\frac{\pi}{2}\right)V^T = \begin{bmatrix} \frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \\ 0 & -1 & 0 \\ \frac{\sqrt{2}}{2} & 0 & -\frac{\sqrt{2}}{2} \end{bmatrix}, \quad UR_Z\left(-\frac{\pi}{2}\right)\Sigma U^T = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix};$$

$$3.\quad UR_Z^T\left(-\frac{\pi}{2}\right)V^T = \begin{bmatrix} \frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \\ 0 & 1 & 0 \\ -\frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \end{bmatrix}, \quad UR_Z\left(-\frac{\pi}{2}\right)\Sigma U^T = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix};$$

$$4.\quad UR_Z^T\left(-\frac{\pi}{2}\right)V^T = \begin{bmatrix} \frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \\ 0 & 1 & 0 \\ -\frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \end{bmatrix}, \quad UR_Z\left(\frac{\pi}{2}\right)\Sigma U^T = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}.$$

Clearly, the third solution is exactly the original motion $(R, \widehat{T})$ except that the translation T is recovered up to a scalar factor (i.e. it is normalized to unit norm). ■

Despite its simplicity, the above algorithm, when used in practice, suffers from some shortcomings that are discussed below.

Number of points

The number of points, eight, assumed by the algorithm, is mostly for convenience and simplicity of presentation. In fact, the matrix E (as a function of (R, T)) has only a total of five degrees of freedom: three for rotation and two for translation (up to a scalar factor). By utilizing some additional algebraic properties of E, we may reduce the necessary number of points. For instance, knowing $\det(E) = 0$, we may relax the condition $\text{rank}(\chi) = 8$ to $\text{rank}(\chi) = 7$, and get two solutions E_1^s and $E_2^s \in \mathbb{R}^9$ from the null space of χ. Nevertheless, there is usually only one $\alpha \in \mathbb{R}$ such that

$$\det(E_1 + \alpha E_2) = 0.$$

Therefore, seven points is all we need to have a relatively simpler algorithm. As shown in Exercise 5.13, in fact, a linear algorithm exists for only six points if more complicated algebraic properties of the essential matrix are used. Hence, it should not be a surprise, as shown by [Kruppa, 1913], that one needs only five points in general position to recover (R, T). It can be shown that there are up to ten (possibly complex) solutions, though the solutions are not obtainable in closed form. Furthermore, for many special motions, one needs only up to four points to determine the associated essential matrix. For instance, planar motions (Exercise 5.6) and motions induced from symmetry (Chapter 10) have this nice property.

Number of solutions and positive depth constraint

Since both E and $-E$ satisfy the same set of epipolar constraints, they in general give rise to $2 \times 2 = 4$ possible solutions for (R, T). However, this does not pose a problem, because only one of the solutions guarantees that the depths of all the 3-D points reconstructed are *positive* with respect to both camera frames. That is, in general, three out of the four solutions will be physically impossible and hence may be discarded (see Exercise 5.11).

Structure requirement: general position

In order for the above algorithm to work properly, the condition that the given eight points be in "general position" is very important. It can be easily shown that if these points form certain degenerate configurations, called critical surfaces, the algorithm will fail (see Exercise 5.14). A case of some practical importance occurs when all the points happen to lie on the same 2-D plane in $\mathbb{R}^3$. We will discuss the geometry for the planar case in Section 5.3, and also later within the context of multiple-view geometry (Chapter 9).

Motion requirement: sufficient parallax

In the derivation of the epipolar constraint we have implicitly assumed that $E \neq 0$, which allowed us to derive the eight-point algorithm where the essential matrix is normalized to $\|E\| = 1$. Due to the structure of the essential matrix, $E = 0 \Leftrightarrow T = 0$. Therefore, the eight-point algorithm requires that the translation (or baseline) $T \neq 0$. The translation T induces parallax in the image plane. In practice, due to noise, the algorithm will likely return an answer even when there is no translation. However, in this case the estimated direction of translation will be meaningless. Therefore, one needs to exercise caution to make sure that there is "sufficient parallax" for the algorithm to be well conditioned. It has been observed experimentally that even for purely rotational motion, i.e. $T = 0$, the "spurious" translation created by noise in the image measurements is sufficient for the eight-point algorithm to return a correct estimate of R.

Infinitesimal viewpoint change

It is often the case in applications that the two views described in this chapter are taken by a moving camera rather than by two static cameras. The derivation of the epipolar constraint and the associated eight-point algorithm does not change, as long as the two vantage points are distinct. In the limit that the two viewpoints come infinitesimally close, the epipolar constraint takes a related but different form called the continuous epipolar constraint, which we will study in Section 5.4. The continuous case is typically of more significance for applications in robot vision, where one is often interested in recovering the linear and angular velocities of the camera.

Multiple motion hypotheses

In the case of multiple moving objects in the scene, image points may no longer satisfy the same epipolar constraint. For example, if we know that there are two independent moving objects with motions, say (R^1, T^1) and (R^2, T^2), then the two images $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ of a point p on one of these objects should satisfy instead the equation

$$(\boldsymbol{x}_2^T E^1 \boldsymbol{x}_1)(\boldsymbol{x}_2^T E^2 \boldsymbol{x}_1) = 0, \tag{5.16}$$

corresponding to the fact that the point p moves according to either motion 1 or motion 2. Here $E^1 = \widehat{T^1} R^1$ and $E^2 = \widehat{T^2} R^2$. As we will see, from this equation it is still possible to recover E^1 and E^2 if enough points are visible on either object. Generalizing to more than two independent motions requires some attention; we will study the multiple-motion problem in Chapter 7.

5.2.2 Euclidean constraints and structure reconstruction

The eight-point algorithm just described uses as input a set of eight or more point correspondences and returns the relative pose (rotation and translation) between the two cameras up to an arbitrary scale $\gamma \in \mathbb{R}^+$. Without loss of generality, we may assume this scale to be $\gamma = 1$, which is equivalent to scaling translation to unit length. Relative pose and point correspondences can then be used to retrieve the position of the points in 3-D by recovering their depths relative to each camera frame.

Consider the basic rigid-body equation, where the pose (R, T) has been recovered, with the translation T defined up to the scale γ. In terms of the images and the depths, it is given by

$$\lambda_2^j \boldsymbol{x}_2^j = \lambda_1^j R \boldsymbol{x}_1^j + \gamma T, \quad j = 1, 2, \ldots, n. \tag{5.17}$$

Notice that since (R, T) are known, the equations given by (5.17) are linear in both the structural scale λ's and the motion scale γ's, and therefore they can be easily solved. For each point, λ_1, λ_2 are its depths with respect to the first and second camera frames, respectively. One of them is therefore redundant; for instance, if λ_1 is known, λ_2 is simply a function of (R, T). Hence we can eliminate, say, λ_2 from the above equation by multiplying both sides by $\widehat{\boldsymbol{x}_2}$, which yields

$$\lambda_1^j \widehat{\boldsymbol{x}_2^j} R \boldsymbol{x}_1^j + \gamma \widehat{\boldsymbol{x}_2^j} T = 0, \quad j = 1, 2, \ldots, n. \tag{5.18}$$

This is equivalent to solving the linear equation

$$M^j \bar{\lambda^j} \doteq \begin{bmatrix} \widehat{\boldsymbol{x}_2^j} R \boldsymbol{x}_1^j, & \widehat{\boldsymbol{x}_2^j} T \end{bmatrix} \begin{bmatrix} \lambda_1^j \\ \gamma \end{bmatrix} = 0, \tag{5.19}$$

where $M^j = \begin{bmatrix} \widehat{\boldsymbol{x}_2^j} R \boldsymbol{x}_1^j, & \widehat{\boldsymbol{x}_2^j} T \end{bmatrix} \in \mathbb{R}^{3 \times 2}$ and $\bar{\lambda^j} = [\lambda_1^j, \gamma]^T \in \mathbb{R}^2$, for $j = 1, 2, \ldots, n$. In order to have a unique solution, the matrix M^j needs to be of rank 1. This is not the case only when $\widehat{\boldsymbol{x}_2}T = 0$, i.e. when the point p lies on the line connecting the two optical centers o_1 and o_2.

Notice that all the n equations above share the same γ; we define a vector $\vec{\lambda} = [\lambda_1^1, \lambda_1^2, \ldots, \lambda_1^n, \gamma]^T \in \mathbb{R}^{n+1}$ and a matrix $M \in \mathbb{R}^{3n \times (n+1)}$ as

$$M \doteq \begin{bmatrix} \widehat{\boldsymbol{x}_2^1}R\boldsymbol{x}_1^1 & 0 & 0 & 0 & 0 & \widehat{\boldsymbol{x}_2^1}T \\ 0 & \widehat{\boldsymbol{x}_2^2}R\boldsymbol{x}_1^2 & 0 & 0 & 0 & \widehat{\boldsymbol{x}_2^2}T \\ 0 & 0 & \ddots & 0 & 0 & \vdots \\ 0 & 0 & 0 & \widehat{\boldsymbol{x}_2^{n-1}}R\boldsymbol{x}_1^{n-1} & 0 & \widehat{\boldsymbol{x}_2^{n-1}}T \\ 0 & 0 & 0 & 0 & \widehat{\boldsymbol{x}_2^n}R\boldsymbol{x}_1^n & \widehat{\boldsymbol{x}_2^n}T \end{bmatrix}. \qquad (5.20)$$

Then the equation

$$M\vec{\lambda} = 0 \qquad (5.21)$$

determines all the unknown depths *up to a single universal scale*. The linear least-squares estimate of $\vec{\lambda}$ is simply the eigenvector of M^TM that corresponds to its smallest eigenvalue. Note that this scale ambiguity is intrinsic, since without any prior knowledge about the scene and camera motion, one cannot disambiguate whether the camera moved twice the distance while looking at a scene twice larger but two times further away.

5.2.3 Optimal pose and structure

The eight-point algorithm given in the previous section assumes that *exact* point correspondences are given. In the presence of noise in image correspondences, we have suggested possible ways of estimating the essential matrix by solving a least-squares problem followed by a projection onto the essential space. But in practice, this will not be satisfying in at least two respects:

1. There is no guarantee that the estimated pose (R, T), is as close as possible to the true solution.
2. Even if we were to accept such an (R, T), a noisy image pair, say $(\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2)$, would not necessarily give rise to a consistent 3-D reconstruction, as shown in Figure 5.6.

At this stage of development, we do not want to bring in all the technical details associated with optimal estimation, since they would bury the geometric intuition. We will therefore discuss only the key ideas, and leave the technical details to Appendix 5.A as well as Chapter 11, where we will address more practical issues.

Choice of optimization objectives

Recall from Chapter 3 that a calibrated camera can be described as a plane perpendicular to the z-axis at a distance 1 from the origin; therefore, the coordinates of image points $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are of the form $[x, y, 1]^T \in \mathbb{R}^3$. In practice, we cannot

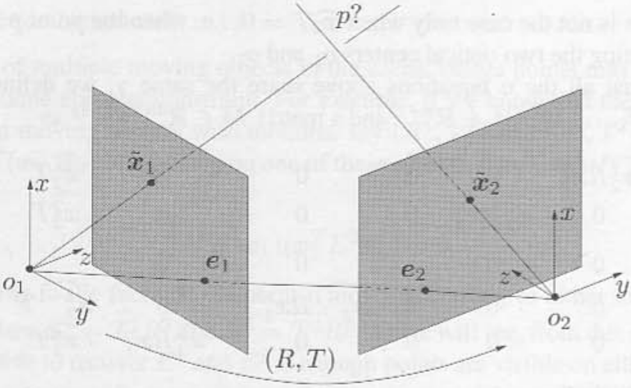

Figure 5.6. Rays extended from a noisy image pair $\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2 \in \mathbb{R}^3$ do not intersect at any point p in 3-D if they do not satisfy the epipolar constraint precisely.

measure the actual coordinates but only their noisy versions, say

$$\tilde{\boldsymbol{x}}_1^j = \boldsymbol{x}_1^j + w_1^j, \quad \tilde{\boldsymbol{x}}_2^j = \boldsymbol{x}_2^j + w_2^j, \quad j = 1, 2, \dots, n, \tag{5.22}$$

where $\boldsymbol{x}_1^j$ and $\boldsymbol{x}_2^j$ denote the "ideal" image coordinates and $w_1^j = [w_{11}^j,\ w_{12}^j,\ 0]^T$ and $w_2^j = [w_{21}^j,\ w_{22}^j,\ 0]^T$ are localization errors in the correspondence. Notice that it is the (unknown) ideal image coordinates $(\boldsymbol{x}_1^j, \boldsymbol{x}_2^j)$ that satisfy the epipolar constraint $\boldsymbol{x}_2^{jT} \widehat{T} R \boldsymbol{x}_1^j = 0$, and *not* the (measured) noisy ones $(\tilde{\boldsymbol{x}}_1^j, \tilde{\boldsymbol{x}}_2^j)$. One could think of the ideal coordinates as a "model," and w_i^j as the discrepancy between the model and the measurements: $\tilde{\boldsymbol{x}}_i^j = \boldsymbol{x}_i^j + w_i^j$. Therefore, in general, we seek the parameters $(\boldsymbol{x}, R, T)$ that minimize the discrepancy between the model and the data, i.e. w_i^j. In order to do so, we first need to decide how to *evaluate* the discrepancy, which determines the choice of optimization objective.

Unfortunately, there is no "correct," uncontroversial, universally accepted objective function, and the choice of discrepancy measure is part of the design process, since it depends on what assumptions are made on the *residuals* w_i^j. Different assumptions result in different choices of discrepancy measures, which eventually result in different "optimal" solutions $(\boldsymbol{x}^*, R^*, T^*)$.

For instance, one may assume that $w = \{w_i^j\}$ are samples from a distribution that depends on the unknown *parameters* $(\boldsymbol{x}, R, T)$, which are considered deterministic but unknown. In this case, based on the model generating the data, one can derive an expression of the *likelihood function* $p(w|\boldsymbol{x}, R, T)$ and choose to maximize it (or, more conveniently, its logarithm) with respect to the unknown parameters. Then the "optimal solution," in the sense of *maximum likelihood*, is given by

$$(\boldsymbol{x}^*, R^*, T^*) = \arg\max \phi_{ML}(\boldsymbol{x}, R, T) \doteq \sum_{i,j} \log p\big((\tilde{\boldsymbol{x}}_i^j - \boldsymbol{x}_i^j)|\boldsymbol{x}, R, T\big).$$

Naturally, different likelihood functions can result in very different optimal solutions. Indeed, there is no guarantee that the maximum is unique, since p can be multimodal, and therefore there may be several choices of parameters that achieve the maximum. Constructing the likelihood function for the location of point features from first principles, starting from the noise characteristics of the photosensitive elements of the sensor, is difficult because of the many nonlinear steps involved in feature detection and tracking. Therefore, it is common to assume that the likelihood belongs to a family of density functions, the most popular choice being the normal (or Gaussian) distribution.

Sometimes, however, one may have reasons to believe that $(\boldsymbol{x}, R, T)$ are not just unknown parameters that can take any value. Instead, even before any measurement is gathered, one can say that some values are more probable than others, a fact that can be described by a joint *a priori* probability density (or *prior*) $p(\boldsymbol{x}, R, T)$. For instance, for a robot navigating on a flat surface, rotation about the horizontal axis may be very improbable, as would translation along the vertical axis. When combined with the likelihood function, the prior can be used to determine the *a posteriori density* , or *posterior* $p(\boldsymbol{x}, R, T|\{\tilde{\boldsymbol{x}}_i^j\})$ using Bayes rule. In this case, one may seek the maximum of the posterior *given* the value of the measurements. This is the *maximum a posteriori* estimate

$$(\boldsymbol{x}^*, R^*, T^*) = \arg\max \phi_{MAP}(\boldsymbol{x}, R, T) \doteq p(\boldsymbol{x}, R, T|\{\tilde{\boldsymbol{x}}_i^j\}).$$

Although this choice has several advantages, in our case it requires defining a probability density on the space of camera poses $SO(3) \times \mathbb{S}^2$, which has a nontrivial geometric structure. This is well beyond the scope of this book, and we will therefore not discuss this criterion further here.

In what follows, we will take a more minimalistic approach to optimality, and simply assume that $\{w_i^j\}$ are unknown values ("errors," or "residuals") whose norms need to be minimized. In this case, we do not postulate any probabilistic description, and we simply seek $(\boldsymbol{x}^*, R^*, T^*) = \arg\min \phi(\boldsymbol{x}, R, T)$, where ϕ is, for instance, the squared 2-norm:

$$\phi(\boldsymbol{x}, R, T) \doteq \sum_j \|w_1^j\|_2^2 + \|w_2^j\|_2^2 = \sum_j \|\tilde{\boldsymbol{x}}_1^j - \boldsymbol{x}_1^j\|_2^2 + \|\tilde{\boldsymbol{x}}_2^j - \boldsymbol{x}_2^j\|_2^2.$$

This corresponds to a *least-squares estimator*. Since $\boldsymbol{x}_1^j$ and $\boldsymbol{x}_2^j$ are the recovered 3-D points projected back onto the image planes, the above criterion is often called the "reprojection error."

However, the unknowns for the above minimization problem are not completely free; for example, they need to satisfy the epipolar constraint $\boldsymbol{x}_2^T \widehat{T} R \boldsymbol{x}_1 = 0$. Hence, with the choice of the least-squares criterion, we can pose the problem of reconstruction as a constrained optimization: given $\tilde{\boldsymbol{x}}_i^j, i = 1, 2,\ j = 1, 2, \ldots, n$, minimize

$$\phi(\boldsymbol{x}, R, T) \doteq \sum_{j=1}^{n} \sum_{i=1}^{2} \|\tilde{\boldsymbol{x}}_i^j - \boldsymbol{x}_i^j\|_2^2 \qquad (5.23)$$

subject to

$$\boldsymbol{x}_2^{jT} \widehat{T} R \boldsymbol{x}_1^j = 0, \quad \boldsymbol{x}_1^{jT} e_3 = 1, \quad \boldsymbol{x}_2^{jT} e_3 = 1, \quad j = 1, 2, \ldots, n. \qquad (5.24)$$

Using Lagrange multipliers (Appendix C), we can convert this constrained optimization problem to an unconstrained one. Details on how to carry out the optimization are outlined in Appendix 5.A.

Remark 5.12 (Equivalence to bundle adjustment). *The reader may have noticed that the depth parameters λ_i, despite being unknown, are missing from the optimization problem of equation (5.24). This is not an oversight: indeed, the depth parameters play the role of Lagrange multipliers in the constrained optimization problem described above, and therefore they enter the optimization problem indirectly. Alternatively, one can write the optimization problem in unconstrained form:*

$$\sum_{j=1}^{n} \left\|\tilde{\boldsymbol{x}}_1^j - \pi_1(\boldsymbol{X}^j)\right\|_2^2 + \left\|\tilde{\boldsymbol{x}}_2^j - \pi_2(\boldsymbol{X}^j)\right\|_2^2, \tag{5.25}$$

where π_1 and π_2 denote the projection of a point $\boldsymbol{X}$ in space onto the first and second images, respectively. If we choose the first camera frame as the reference, then the above expression can be simplified to[6]

$$\phi(\boldsymbol{x}_1, R, T, \lambda) = \sum_{j=1}^{n} \left\|\tilde{\boldsymbol{x}}_1^j - \boldsymbol{x}_1^j\right\|_2^2 + \left\|\tilde{\boldsymbol{x}}_2^j - \pi(R\lambda_1^j \boldsymbol{x}_1^j + T)\right\|_2^2. \tag{5.26}$$

Minimizing the above expression with respect to the unknowns $(R, T, \boldsymbol{x}_1, \lambda)$ is known in the literature as bundle adjustment. *Bundle adjustment and the constrained optimization described above are simply two different ways to parameterize the same optimization objective. As we will see in Appendix 5.A, the constrained form better highlights the geometric structure of the problem, and serves as a guide to develop effective approximations.*

In the remainder of this section, we limit ourselves to describing a simplified cost functional that approximates the reprojection error resulting in simpler optimization algorithms, while retaining a strong geometric interpretation. In this approximation, the unknown $\boldsymbol{x}$ is approximated by the measured $\tilde{\boldsymbol{x}}$, so that the cost function ϕ depends only on camera pose (R, T) (see Appendix 5.A for more details):

$$\phi(R, T) \doteq \sum_{j=1}^{n} \frac{(\tilde{\boldsymbol{x}}_2^{jT} \widehat{T} R \tilde{\boldsymbol{x}}_1^j)^2}{\|\widehat{e}_3 \widehat{T} R \tilde{\boldsymbol{x}}_1^j\|^2} + \frac{(\tilde{\boldsymbol{x}}_2^{jT} \widehat{T} R \tilde{\boldsymbol{x}}_1^j)^2}{\|\tilde{\boldsymbol{x}}_2^{jT} \widehat{T} R \widehat{e}_3^T\|^2}. \tag{5.27}$$

Geometrically, this expression can be interpreted as distances from the image points $\tilde{\boldsymbol{x}}_1^j$ and $\tilde{\boldsymbol{x}}_2^j$ to corresponding epipolar lines in the two image planes, respectively, as shown in Figure 5.7. For instance, the reader can verify as an exercise

[6] Here we use π to denote the standard planar projection introduced in Chapter 3: $[X, Y, Z]^T \mapsto [X/Z, Y/Z, 1]^T$.

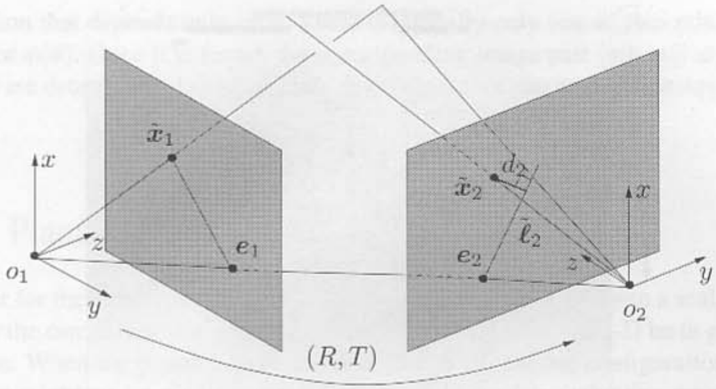

Figure 5.7. Two noisy image points $\tilde{x}_1, \tilde{x}_2 \in \mathbb{R}^3$. Here $\tilde{\ell}_2$ is an epipolar line that is the intersection of the second image plane with the epipolar plane. The distance d_2 is the geometric distance between the second image point $\tilde{x}_2$ and the epipolar line. Symmetrically, one can define a similar geometric distance d_1 in the first image plane.

(Exercise 5.12) that following the notation in the figure, we have

$$d_2^2 = \frac{(\tilde{x}_2^T \widehat{T} R \tilde{x}_1)^2}{\|\widehat{e}_3 \widehat{T} R \tilde{x}_1\|^2}.$$

In the presence of noise, minimizing the above objective function, although more difficult, improves the results of the linear eight-point algorithm.

Example 5.13 (Comparison with the linear algorithm). Figure 5.8 demonstrates the effect of the optimization: numerical simulations were run for both the linear eight-point algorithm and the nonlinear optimization. Values of the objective function $\phi(R, T)$ at different T are plotted (with R fixed at the ground truth); "+" denotes the true translation T, "$*$" is the estimated T from the linear eight-point algorithm, and "$\circ$" is the estimated T by upgrading the linear algorithm result with the optimization. ■

Structure triangulation

If we were given the optimal estimate of camera pose (R, T), obtained, for instance, from Algorithm 5.5 in Appendix 5.A, we can find a pair of images (x_1^*, x_2^*) that satisfy the epipolar constraint $x_2^T \widehat{T} R x_1 = 0$ and minimize the (reprojection) error

$$\phi(x) = \|\tilde{x}_1 - x_1\|^2 + \|\tilde{x}_2 - x_2\|^2. \qquad (5.28)$$

This is called the *triangulation problem*. The key to its solution is to find what exactly the reprojection error depends on, which can be more easily explained geometrically by Figure 5.9. As we see from the figure, the value of the reprojection error depends only on the position of the epipolar plane P: when the plane P rotates around the baseline (o_1, o_2), the image pair (x_1, x_2), which minimizes the distance $\|\tilde{x}_1 - x_1\|^2 + \|\tilde{x}_2 - x_2\|^2$, changes accordingly, and so does the error. To

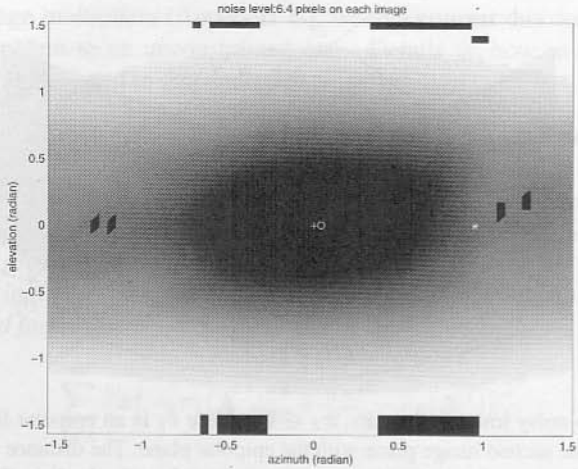

Figure 5.8. Improvement by nonlinear optimization. A two-dimensional projection of the five-dimensional residual function $\phi(R, T)$ is shown in greyscale. The residual corresponds to the two-dimensional function $\phi(\hat{R}, T)$ with rotation fixed at the true value. The location of the solution found by the linear algorithm is shown as "$*$," and it can be seen that it is quite far from the true minimum (darkest point in the center of the image, marked by "+").The solution obtained by nonlinear optimization is marked by "$\circ$," which shows a significant improvement.

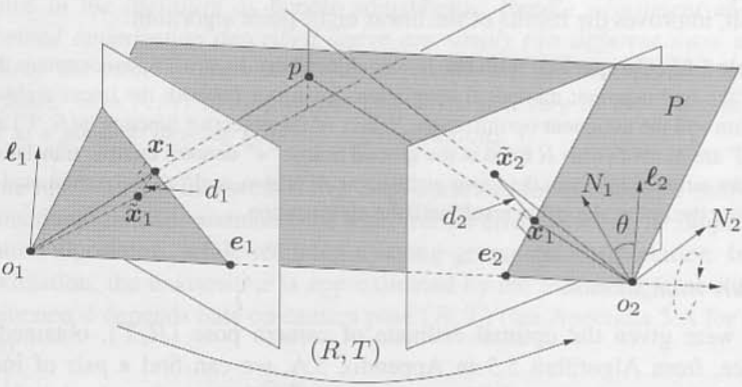

Figure 5.9. For a fixed epipolar plane P, the pair of images $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ that minimize the reprojection error $d_1^2 + d_2^2$ must be points on the two epipolar lines and closest to $\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2$, respectively. Hence the reprojection error is a function only of the position of the epipolar plane P.

parameterize the position of the epipolar plane, let $(\boldsymbol{e}_2, N_1, N_2)$ be an orthonormal basis in the second camera frame. Then P is determined by its normal vector $\boldsymbol{\ell}_2$ (with respect to the second camera frame), which in turn is determined by the angle θ between $\boldsymbol{\ell}_2$ and N_1 (Figure 5.9). Hence the reprojection error ϕ should be a function that depends only on θ. There is typically only one θ^* that minimizes the error $\phi(\theta)$. Once it is found, the corresponding image pair $(\boldsymbol{x}_1^*, \boldsymbol{x}_2^*)$ and 3-D point p are determined. Details of the related algorithm can be found in Appendix 5.A.

5.3 Planar scenes and homography

In order for the eight-point algorithm to give a unique solution (up to a scalar factor) for the camera motion, it is crucial that the feature points in 3-D be in general position. When the points happen to form certain degenerate configurations, the solution might no longer be unique. Exercise 5.14 explains why this may occur when all the feature points happen to lie on certain 2-D surfaces, called critical surfaces.[7] Many of these critical surfaces occur rarely in practice, and their importance is limited. However, 2-D planes, which happen to be a special case of critical surfaces, are ubiquitous in man-made environments and in aerial imaging.

Therefore, if one applies the eight-point algorithm to images of points all lying on the same 2-D plane, the algorithm will fail to provide a unique solution (as we will soon see why). On the other hand, in many applications, a scene can indeed be approximately planar (e.g., the landing pad for a helicopter) or piecewise planar (e.g., the corridors inside a building). We therefore devote this section to this special but important case.

5.3.1 Planar homography

Let us consider two images of points p on a 2-D plane P in 3-D space. For simplicity, we will assume throughout the section that the optical center of the camera never passes through the plane.

Now suppose that two images $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ are given for a point $p \in P$ with respect to two camera frames. Let the coordinate transformation between the two frames be

$$\boldsymbol{X}_2 = R\boldsymbol{X}_1 + T, \tag{5.29}$$

where $\boldsymbol{X}_1, \boldsymbol{X}_2$ are the coordinates of p relative to camera frames 1 and 2, respectively. As we have already seen, the two images $\boldsymbol{x}_1, \boldsymbol{x}_2$ of p satisfy the epipolar constraint

$$\boldsymbol{x}_2^T E \boldsymbol{x}_1 = \boldsymbol{x}_2^T \widehat{T} R \boldsymbol{x}_1 = 0.$$

However, for points on the same plane P, their images will share an extra constraint that makes the epipolar constraint alone no longer sufficient.

[7] In general, such critical surfaces can be described by certain quadratic equations in the X, Y, Z coordinates of the point, hence are often referred to as quadratic surfaces.

Let $N = [n_1, n_2, n_2]^T \in \mathbb{S}^2$ be the unit normal vector of the plane P with respect to the first camera frame, and let $d > 0$ denote the distance from the plane P to the optical center of the first camera. Then we have

$$N^T \boldsymbol{X}_1 = n_1 X + n_2 Y + n_3 Z = d \quad \Leftrightarrow \quad \frac{1}{d} N^T \boldsymbol{X}_1 = 1, \quad \forall \boldsymbol{X}_1 \in P. \tag{5.30}$$

Substituting equation (5.30) into equation (5.29) gives

$$\boldsymbol{X}_2 = R\boldsymbol{X}_1 + T = R\boldsymbol{X}_1 + T\frac{1}{d}N^T\boldsymbol{X}_1 = \left(R + \frac{1}{d}TN^T\right)\boldsymbol{X}_1. \tag{5.31}$$

We call the matrix

$$H \doteq R + \frac{1}{d}TN^T \in \mathbb{R}^{3\times 3} \tag{5.32}$$

the *(planar) homography matrix*, since it denotes a linear transformation from $\boldsymbol{X}_1 \in \mathbb{R}^3$ to $\boldsymbol{X}_2 \in \mathbb{R}^3$ as

$$\boldsymbol{X}_2 = H\boldsymbol{X}_1.$$

Note that the matrix H depends on the motion parameters $\{R, T\}$ as well as the structure parameters $\{N, d\}$ of the plane P. Due to the inherent scale ambiguity in the term $\frac{1}{d}T$ in equation (5.32), one can at most expect to recover from H the ratio of the translation T scaled by the distance d. From

$$\lambda_1 \boldsymbol{x}_1 = \boldsymbol{X}_1, \quad \lambda_2 \boldsymbol{x}_2 = \boldsymbol{X}_2, \quad \boldsymbol{X}_2 = H\boldsymbol{X}_1, \tag{5.33}$$

we have

$$\lambda_2 \boldsymbol{x}_2 = H\lambda_1 \boldsymbol{x}_1 \quad \Leftrightarrow \quad \boldsymbol{x}_2 \sim H\boldsymbol{x}_1, \tag{5.34}$$

where we recall that $\sim$ indicates equality up to a scalar factor. Often, the equation

$$\boxed{\boldsymbol{x}_2 \sim H\boldsymbol{x}_1} \tag{5.35}$$

itself is referred to as a *(planar) homography* mapping induced by a plane P. Despite the scale ambiguity, as illustrated in Figure 5.10, H introduces a special map between points in the first image and those in the second in the following sense:

1. For any point $\boldsymbol{x}_1$ in the first image that is the image of some point, say p on the plane P, its corresponding second image $\boldsymbol{x}_2$ is uniquely determined as $\boldsymbol{x}_2 \sim H\boldsymbol{x}_1$, since for any other point, say $\boldsymbol{x}_2'$, on the same epipolar line $\boldsymbol{\ell}_2 \sim E\boldsymbol{x}_1 \in \mathbb{R}^3$, the ray $o_2\boldsymbol{x}_2'$ will intersect the ray $o_1\boldsymbol{x}_1$ at a point p' out of the plane.

2. On the other hand, if $\boldsymbol{x}_1$ is the image of some point, say p', not on the plane P, then $\boldsymbol{x}_2 \sim H\boldsymbol{x}_1$ is only a point that is on the same epipolar line $\boldsymbol{\ell}_2 \sim E\boldsymbol{x}_1$ as its actual corresponding image $\boldsymbol{x}_2'$. That is, $\boldsymbol{\ell}_2^T\boldsymbol{x}_2 = \boldsymbol{\ell}_2^T\boldsymbol{x}_2' = 0$.

We hence have the following result:

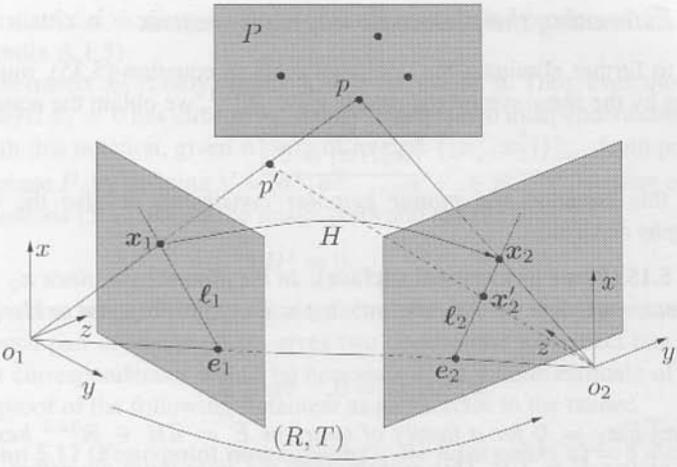

Figure 5.10. Two images $x_1, x_2 \in \mathbb{R}^3$ of a 3-D point p on a plane P. They are related by a homography H that is induced by the plane.

Proposition 5.14 (Homography for epipolar lines). *Given a homography H (induced by plane P in 3-D) between two images, for any pair of corresponding images (x_1, x_2) of a 3-D point p that is not necessarily on P, the associated epipolar lines are*

$$\boxed{\ell_2 \sim \widehat{x_2} H x_1, \quad \ell_1 \sim H^T \ell_2.} \tag{5.36}$$

Proof. If p is not on P, the first equation is true from point 2 in above discussion. Note that for points on the plane P, $x_2 = Hx_1$ implies $\widehat{x_2}Hx_1 = 0$, and the first equation is still true as long as we adopt the convention that $v \sim 0$, $\forall v \in \mathbb{R}^3$. The second equation is easily proven using the definition of a line $\ell^T x = 0$. □

This property of the homography allows one to compute epipolar lines without knowing the essential matrix. We will explore further the relationships between the essential matrix and the planar homography in Section 5.3.4.

In addition to the fact that the homography matrix H encodes information about the camera motion and the scene structure, knowing it directly facilitates establishing correspondence between points in the first and the second images. As we will see soon, H can be computed in general from a small number of corresponding image pairs. Once H is known, correspondence between images of other points on the same plane can then be fully established, since the corresponding location x_2 for an image point x_1 is simply Hx_1. Proposition 5.14 suggests that correspondence between images of points not on the plane can also be established, since H contains information about the epipolar lines.

5.3.2 Estimating the planar homography matrix

In order to further eliminate the unknown scale in equation (5.35), multiplying both sides by the skew-symmetric matrix $\widehat{x_2} \in \mathbb{R}^{3\times 3}$, we obtain the equation

$$\boxed{\widehat{x_2} H x_1 = 0.} \tag{5.37}$$

We call this equation the *planar epipolar constraint*, or also the *(planar) homography constraint*.

Remark 5.15 (Plane as a critical surface). *In the planar case, since $x_2 \sim Hx_1$, for any vector $u \in \mathbb{R}^3$, we have that $u \times x_2 = \widehat{u}x_2$ is orthogonal to Hx_1. Hence we have*

$$x_2^T \widehat{u} H x_1 = 0, \quad \forall u \in \mathbb{R}^3.$$

That is, $x_2^T E x_1 = 0$ for a family of matrices $E = \widehat{u}H \in \mathbb{R}^{3\times 3}$ besides the essential matrix $E = \widehat{T}R$. This explains why the eight-point algorithm does not apply to feature points from a planar scene.

Example 5.16 (Homography from a pure rotation). The homographic relation $x_2 \sim Hx_1$ also shows up when the camera is purely rotating, i.e. $X_2 = RX_1$. In this case, the homography matrix H becomes $H = R$, since $T = 0$. Consequently, we have the constraint

$$\widehat{x_2} R x_1 = 0.$$

One may view this as a special planar scene case, since without translation, information about the depth of the scene is completely lost in the images, and one might as well interpret the scene to be planar (e.g., all the points lie on a plane infinitely far away). As the distance of the plane d goes to infinity, $\lim_{d\to\infty} H = R$.

The homography from purely rotational motion can be used to construct image mosaics of the type shown in Figure 5.11. For additional references on how to construct panoramic mosaics the reader can refer to [Szeliski and Shum, 1997, Sawhney and Kumar, 1999], where the latter includes compensation for radial distortion. ■

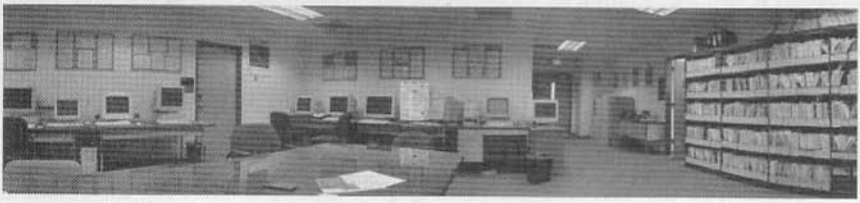

Figure 5.11. Mosaic from the rotational homography.

Since equation (5.37) is *linear* in H, by stacking the entries of H as a vector,

$$H^s \doteq [H_{11}, H_{21}, H_{31}, H_{12}, H_{22}, H_{32}, H_{13}, H_{23}, H_{33}]^T \quad \in \mathbb{R}^9, \tag{5.38}$$

we may rewrite equation (5.37) as

$$a^T H^s = 0,$$

where the matrix $\boldsymbol{a} \doteq \boldsymbol{x}_1 \otimes \widehat{\boldsymbol{x}_2} \in \mathbb{R}^{9\times 3}$ is the Kronecker product of $\widehat{\boldsymbol{x}_2}$ and $\boldsymbol{x}_1$ (see Appendix A.1.3).

Since the matrix $\widehat{\boldsymbol{x}_2}$ is only of rank 2, so is the matrix $\boldsymbol{a}$. Thus, even though the equation $\widehat{\boldsymbol{x}_2}H\boldsymbol{x}_1 = 0$ has three rows, it only imposes two independent constraints on H. With this notation, given n pairs of images $\{(\boldsymbol{x}_1^j, \boldsymbol{x}_2^j)\}_{j=1}^n$ from points on the same plane P, by defining $\chi \doteq [\boldsymbol{a}^1, \boldsymbol{a}^2, \ldots, \boldsymbol{a}^n]^T \in \mathbb{R}^{3n\times 9}$, we may combine all the equations (5.37) for all the image pairs and rewrite them as

$$\chi H^s = 0. \tag{5.39}$$

In order to solve uniquely (up to a scalar factor) for H^s, we must have $\text{rank}(\chi) = 8$. Since each pair of image points gives two constraints, we expect that at least four point correspondences would be necessary for a unique estimate of H. We leave the proof of the following statement as an exercise to the reader.

Proposition 5.17 (Four-point homography). *We have $rank(\chi) = 8$ if and only if there exists a set of four points (out of the n) such that no three of them are collinear; i.e. they are in a general configuration in the plane.*

Thus, if there are more than four image correspondences of which no three in each image are collinear, we may apply standard linear least-squares estimation to find $\min \|\chi H^s\|^2$ to recover H up to a scalar factor. That is, we are able to recover H of the form

$$H_L \doteq \lambda H = \lambda\left(R + \frac{1}{d}TN^T\right) \quad \in \mathbb{R}^{3\times 3} \tag{5.40}$$

for some (unknown) scalar factor λ.

Knowing H_L, the next thing is obviously to determine the scalar factor λ by taking into account the structure of H.

Lemma 5.18 (Normalization of the planar homography). *For a matrix of the form $H_L = \lambda\left(R + \frac{1}{d}TN^T\right)$, we have*

$$|\lambda| = \sigma_2(H_L), \tag{5.41}$$

where $\sigma_2(H_L) \in \mathbb{R}$ is the second largest singular value of H_L.

Proof. Let $u = \frac{1}{d}R^T T \in \mathbb{R}^3$. Then we have

$$H_L^T H_L = \lambda^2(I + uN^T + Nu^T + \|u\|^2 NN^T).$$

Obviously, the vector $u \times N = \widehat{u}N \in \mathbb{R}^3$, which is orthogonal to both u and N, is an eigenvector and $H_L^T H_L(\widehat{u}N) = \lambda^2(\widehat{u}N)$. Hence $|\lambda|$ is a singular value of H_L. We only have to show that it is the second largest. Let $v = \|u\|N, w = u/\|u\| \in \mathbb{R}^3$. We have

$$Q = uN^T + Nu^T + \|u\|^2 NN^T = (w+v)(w+v)^T - ww^T.$$

The matrix Q has a positive, a negative, and a zero eigenvalue, except that when $u \sim N$, Q will have two repeated zero eigenvalues. In any case, $H_L^T H_L$ has λ^2 as its second-largest eigenvalue. □

Then, if $\{\sigma_1, \sigma_2, \sigma_3\}$ are the singular values of H_L recovered from linear least-squares estimation, we set a new

$$H = H_L / \sigma_2(H_L).$$

This recovers H up to the form $H = \pm\left(R + \frac{1}{d}TN^T\right)$. To get the correct sign, we may use $\lambda_2^j \boldsymbol{x}_2^j = H\lambda_1^j \boldsymbol{x}_1^j$ and the fact that $\lambda_1^j, \lambda_2^j > 0$ to impose the positive depth constraint

$$(\boldsymbol{x}_2^j)^T H \boldsymbol{x}_1^j > 0, \quad \forall j = 1, 2, \ldots, n.$$

Thus, if the points $\{p\}_{j=1}^n$ are in general configuration on the plane, then the matrix $H = \left(R + \frac{1}{d}TN^T\right)$ can be uniquely determined from the image pair.

5.3.3 Decomposing the planar homography matrix

After we have recovered H of the form $H = \left(R + \frac{1}{d}TN^T\right)$, we now study how to decompose such a matrix into its motion and structure parameters, namely $\{R, \frac{T}{d}, N\}$.

Theorem 5.19 (Decomposition of the planar homography matrix). *Given a matrix $H = \left(R + \frac{1}{d}TN^T\right)$, there are at most two physically possible solutions for a decomposition into parameters $\left\{R, \frac{1}{d}T, N\right\}$ given in Table 5.1.*

Proof. First notice that H preserves the length of any vector orthogonal to N, i.e. if $N \perp a$ for some $a \in \mathbb{R}^3$, we have $\|Ha\|^2 = \|Ra\|^2 = \|a\|^2$. Also, if we know the plane spanned by the vectors that are orthogonal to N, we then know N itself. Let us first recover the vector N based on this knowledge.

The symmetric matrix $H^T H$ will have three eigenvalues $\sigma_1^2 \geq \sigma_2^2 \geq \sigma_3^2 \geq 0$, and from Lemma 5.18 we know that $\sigma_2 = 1$. Since $H^T H$ is symmetric, it can be diagonalized by an orthogonal matrix $V \in SO(3)$ such that

$$H^T H = V \Sigma V^T, \tag{5.42}$$

where $\Sigma = \operatorname{diag}\{\sigma_1^2, \sigma_2^2, \sigma_3^2\}$. If $[v_1, v_2, v_3]$ are the three column vectors of V, we have

$$H^T H v_1 = \sigma_1^2 v_1, \quad H^T H v_2 = v_2, \quad H^T H v_3 = \sigma_3^2 v_3. \tag{5.43}$$

Hence v_2 is orthogonal to both N and T, and its length is preserved under the map H. Also, it is easy to check that the length of two other unit-length vectors defined as

$$u_1 \doteq \frac{\sqrt{1-\sigma_3^2}v_1 + \sqrt{\sigma_1^2 - 1}v_3}{\sqrt{\sigma_1^2 - \sigma_3^2}}, \quad u_2 \doteq \frac{\sqrt{1-\sigma_3^2}v_1 - \sqrt{\sigma_1^2 - 1}v_3}{\sqrt{\sigma_1^2 - \sigma_3^2}} \tag{5.44}$$

is also preserved under the map H. Furthermore, it is easy to verify that H preserves the length of any vectors inside each of the two subspaces

$$S_1 = \operatorname{span}\{v_2, u_1\}, \quad S_2 = \operatorname{span}\{v_2, u_2\}. \tag{5.45}$$

Since v_2 is orthogonal to u_1 and u_2, $\widehat{v_2}u_1$ is a unit normal vector to S_1, and $\widehat{v_2}u_2$ a unit normal vector to S_2. Then $\{v_2, u_1, \widehat{v_2}u_1\}$ and $\{v_2, u_2, \widehat{v_2}u_2\}$ form two sets of orthonormal bases for $\mathbb{R}^3$. Notice that we have

$$Rv_2 = Hv_2, \quad Ru_i = Hu_i, \quad R(\widehat{v_2}u_i) = \widehat{Hv_2}Hu_i$$

if N is the normal to the subspace $S_i, i = 1, 2$, as shown in Figure 5.12.

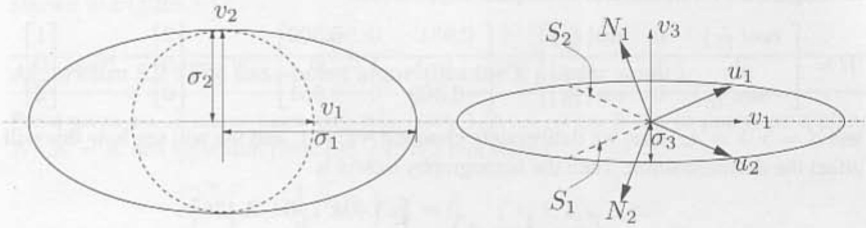

Figure 5.12. In terms of singular vectors (v_1, v_2, v_3) and singular values $(\sigma_1, \sigma_2, \sigma_3)$ of the matrix H, there are two candidate subspaces S_1 and S_2 on which the vectors' length is preserved by the homography matrix H.

Define the matrices

$$U_1 = [v_2, u_1, \widehat{v_2}u_1], \quad W_1 = [Hv_2, Hu_1, \widehat{Hv_2}Hu_1];$$
$$U_2 = [v_2, u_2, \widehat{v_2}u_2], \quad W_2 = [Hv_2, Hu_2, \widehat{Hv_2}Hu_2].$$

We then have

$$RU_1 = W_1, \quad RU_2 = W_2.$$

This suggests that each subspace S_1, or S_2 may give rise to a solution to the decomposition. By taking into account the extra sign ambiguity in the term $\frac{1}{d}TN^T$, we then obtain four solutions for decomposing $H = R + \frac{1}{d}TN^T$ to $\{R, \frac{1}{d}T, N\}$. They are given in Table 5.1.

Solution 1	$R_1 = W_1U_1^T$ $N_1 = \widehat{v_2}u_1$ $\frac{1}{d}T_1 = (H - R_1)N_1$	Solution 3	$R_3 = R_1$ $N_3 = -N_1$ $\frac{1}{d}T_3 = -\frac{1}{d}T_1$
Solution 2	$R_2 = W_2U_2^T$ $N_2 = \widehat{v_2}u_2$ $\frac{1}{d}T_2 = (H - R_2)N_2$	Solution 4	$R_4 = R_2$ $N_4 = -N_2$ $\frac{1}{d}T_4 = -\frac{1}{d}T_2$

Table 5.1. Four solutions for the planar homography decomposition, only two of which satisfy the positive depth constraint.

In order to reduce the number of physically possible solutions, we may impose the positive depth constraint (Exercise 5.11); since the camera can see only points that are in front of it, we must have $N^Te_3 = n_3 > 0$. Suppose that solution 1 is the true one; this constraint will then eliminate solution 3 as being physically impossible. Similarly, one of solutions 2 or 4 will be eliminated. For the case that $T \sim N$, we have $\sigma_3^2 = 0$ in the above proof. Hence $u_1 = u_2$, and solutions 1 and 2 are equivalent. Imposing the positive depth constraint leads to a unique solution for all motion and structure parameters. □

Example 5.20 (A numerical example). Suppose that

$$R = \begin{bmatrix} \cos(\frac{\pi}{10}) & 0 & \sin(\frac{\pi}{10}) \\ 0 & 1 & 0 \\ -\sin(\frac{\pi}{10}) & 0 & \cos(\frac{\pi}{10}) \end{bmatrix} = \begin{bmatrix} 0.951 & 0 & 0.309 \\ 0 & 1 & 0 \\ -0.309 & 0 & 0.951 \end{bmatrix}, \quad T = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}, \quad N = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix},$$

and $d = 5, \lambda = 4$. Here, we deliberately choose $\|N\| \neq 1$, and we will see how this will affect the decomposition. Then the homography matrix is

$$H_L = \lambda \left(R + \frac{1}{d} T N^T \right) = \begin{bmatrix} 5.404 & 0 & 4.436 \\ 0 & 4 & 0 \\ -1.236 & 0 & 3.804 \end{bmatrix}.$$

The singular values of H_L are $\{7.197, 4.000, 3.619\}$. The middle one is exactly the scale λ. Hence for the normalized homography matrix $H_L/4 \rightarrow H$, the matrix $H^T H$ has the SVD[8]

$$V \Sigma V^T \doteq \begin{bmatrix} 0.675 & 0 & -0.738 \\ 0 & 1 & 0 \\ 0.738 & 0 & 0.675 \end{bmatrix} \begin{bmatrix} 3.237 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0.819 \end{bmatrix} \begin{bmatrix} 0.675 & 0 & -0.738 \\ 0 & 1 & 0 \\ 0.738 & 0 & 0.675 \end{bmatrix}^T.$$

Then the two vectors u_1 and u_2 are given by

$$u_1 = [-0.525, 0, 0.851]^T; \quad u_2 = [0.894, 0, -0.447]^T.$$

The four solutions to the decomposition are

$$R_1 = \begin{bmatrix} 0.704 & 0 & 0.710 \\ 0 & 1 & 0 \\ -0.710 & 0 & 0.704 \end{bmatrix}, \quad N_1 = \begin{bmatrix} 0.851 \\ 0 \\ 0.525 \end{bmatrix}, \quad \frac{1}{d} T_1 = \begin{bmatrix} 0.760 \\ 0 \\ 0.471 \end{bmatrix};$$

$$R_2 = \begin{bmatrix} 0.951 & 0 & 0.309 \\ 0 & 1 & 0 \\ -0.309 & 0 & 0.951 \end{bmatrix}, \quad N_2 = \begin{bmatrix} -0.447 \\ 0 \\ -0.894 \end{bmatrix}, \quad \frac{1}{d} T_2 = \begin{bmatrix} -0.894 \\ 0 \\ 0 \end{bmatrix};$$

$$R_3 = \begin{bmatrix} 0.704 & 0 & 0.710 \\ 0 & 1 & 0 \\ -0.710 & 0 & 0.704 \end{bmatrix}, \quad N_3 = \begin{bmatrix} -0.851 \\ 0 \\ -0.525 \end{bmatrix}, \quad \frac{1}{d} T_3 = \begin{bmatrix} -0.760 \\ 0 \\ -0.471 \end{bmatrix};$$

$$R_4 = \begin{bmatrix} 0.951 & 0 & 0.309 \\ 0 & 1 & 0 \\ -0.309 & 0 & 0.951 \end{bmatrix}, \quad N_4 = \begin{bmatrix} 0.447 \\ 0 \\ 0.894 \end{bmatrix}, \quad \frac{1}{d} T_4 = \begin{bmatrix} 0.894 \\ 0 \\ 0 \end{bmatrix}.$$

Obviously, the fourth solution is the correct one: The original $\|N\| \neq 1$, and N is recovered up to a scalar factor (with its length normalized to 1), and hence in the solution we should expect $\frac{1}{d} T_4 = \frac{\|N\|}{d} T$. Notice that the first solution also satisfies $N_1^T e_3 > 0$,

[8] The Matlab routine SVD does not always guarantee that $V \in SO(3)$. When using the routine, if one finds that $\det(V) = -1$, replace both V's by $-V$.

which indicates a plane in front of the camera. Hence it corresponds to another physically possible solution (from the decomposition). ■

We will investigate the geometric relation between the remaining two physically possible solutions in the exercises (see Exercise 5.19). We conclude this section by presenting the following four-point Algorithm 5.2 for motion estimation from a planar scene. Examples of use of this algorithm on real images are shown in Figure 5.13.

Algorithm 5.2 (The four-point algorithm for a planar scene).

For a given set of image pairs $(\boldsymbol{x}_1^j, \boldsymbol{x}_2^j)$, $j = 1, 2, \ldots, n$ $(n \geq 4)$, of points on a plane $N^T \boldsymbol{X} = d$, this algorithm finds $\{R, \frac{1}{d}T, N\}$ that solves

$$\widehat{\boldsymbol{x}_2^j}^T \left(R + \frac{1}{d} T N^T \right) \boldsymbol{x}_1^j = 0, \quad j = 1, 2, \ldots, n.$$

1. **Compute a first approximation of the homography matrix**
   Construct $\mathcal{X} = [\boldsymbol{a}^1, \boldsymbol{a}^2, \ldots, \boldsymbol{a}^n]^T \in \mathbb{R}^{3n \times 9}$ from correspondences $\boldsymbol{x}_1^j$ and $\boldsymbol{x}_2^j$, where $\boldsymbol{a}^j = \boldsymbol{x}_1^j \otimes \widehat{\boldsymbol{x}_2^j} \in \mathbb{R}^{9 \times 3}$. Find the vector $H_L^s \in \mathbb{R}^9$ of unit length that solves
   $$\mathcal{X} H_L^s = 0$$
   as follows: compute the SVD of $\mathcal{X} = U_{\mathcal{X}} \Sigma_{\mathcal{X}} V_{\mathcal{X}}^T$ and define H_L^s to be the ninth column of $V_{\mathcal{X}}$. Unstack the nine elements of H_L^s into a square 3×3 matrix H_L.

2. **Normalization of the homography matrix**
   Compute the eigenvalues $\{\sigma_1, \sigma_2, \sigma_3\}$ of the matrix H_L and normalize it as
   $$H = H_L / \sigma_2.$$
   Correct the sign of H according to sign $\left((\boldsymbol{x}_2^j)^T H \boldsymbol{x}_1^j \right)$ for $j = 1, 2, \ldots, n$.

3. **Decomposition of the homography matrix**
   Compute the singular value decomposition of
   $$H^T H = V \Sigma V^T$$
   and compute the four solutions for a decomposition $\{R, \frac{1}{d}T, N\}$ as in the proof of Theorem 5.19. Select the two physically possible ones by imposing the positive depth constraint $N^T e_3 > 0$.

5.3.4 Relationships between the homography and the essential matrix

In practice, especially when the scene is piecewise planar, we often need to compute the essential matrix E with a given homography H computed from some four points known to be planar; or in the opposite situation, the essential matrix E may have been already estimated using points in general position, and we then want to compute the homography for a particular (usually smaller) set of coplanar

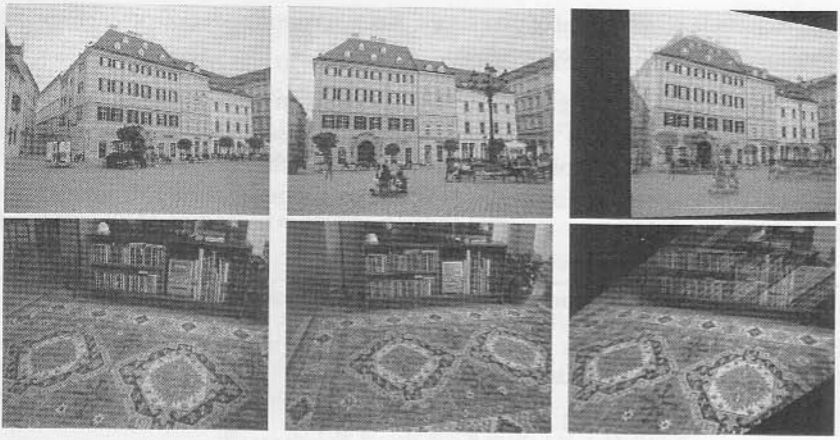

Figure 5.13. Homography between the left and middle images is determined by the building facade on the top, and the ground plane on the bottom. The right image is the warped image overlayed on the first image based on the estimated homography H. Note that all points on the reference plane are aligned, whereas points outside the reference plane are offset by an amount that is proportional to their distance from the reference plane.

points. We hence need to understand the relationship between the essential matrix E and the homography H.

Theorem 5.21 (Relationships between the homography and essential matrix). *For a matrix $E = \widehat{T}R$ and a matrix $H = R + Tu^T$ for some nonsingular $R \in \mathbb{R}^{3\times 3}$, $T, u \in \mathbb{R}^3$, with $\|T\| = 1$, we have:*

1. *$E = \widehat{T}H$;*
2. *$H^T E + E^T H = 0$;*
3. *$H = \widehat{T}^T E + Tv^T$, for some $v \in \mathbb{R}^3$.*

Proof. The proof of item 1 is easy, since $\widehat{T}T = 0$. For item 2, notice that $H^T E = (R + Tu^T)^T \widehat{T}R = R^T \widehat{T}R$ is a skew-symmetric matrix, and hence $H^T E = -E^T H$. For item 3, notice that

$$\widehat{T}H = \widehat{T}R = \widehat{T}\widehat{T}^T\widehat{T}R = \widehat{T}\widehat{T}^T E,$$

since $\widehat{T}\widehat{T}^T v = (I - TT^T)v$ represents an orthogonal projection of v onto the subspace (a plane) orthogonal to T (see Exercise 5.3). Therefore, $\widehat{T}(H - \widehat{T}^T E) = 0$. That is, all the columns of $H - \widehat{T}^T E$ are parallel to T, and hence we have $H - \widehat{T}^T E = Tv^T$ for some $v \in \mathbb{R}^3$. □

Notice that neither the statement nor the proof of the theorem assumes that R is a rotation matrix. Hence, the results will also be applicable to the case in which the camera is not calibrated, which will be discussed in the next chapter.

This theorem directly implies two useful corollaries stated below that allow us to easily compute E from H as well as H from E with minimum extra information from images.[9] The first corollary is a direct consequence of the above theorem and Proposition 5.14:

Corollary 5.22 (From homography to the essential matrix). *Given a homography H and two pairs of images $(\boldsymbol{x}_1^i, \boldsymbol{x}_2^i), i = 1, 2$, of two points not on the plane P from which H is induced, we have*

$$E = \widehat{T}H, \tag{5.46}$$

where $T \sim \widehat{\ell_2^1}\ell_2^2$ and $\|T\| = 1$.

Proof. According to Proposition 5.14, ℓ_2^i is the epipolar line $\ell_2^i \sim \widehat{\boldsymbol{x}_2^i}H\boldsymbol{x}_1^i$, $i = 1, 2$. Both epipolar lines ℓ_2^1, ℓ_2^2 pass through the epipole $e_2 \sim T$. This can be illustrated by Figure 5.14. □

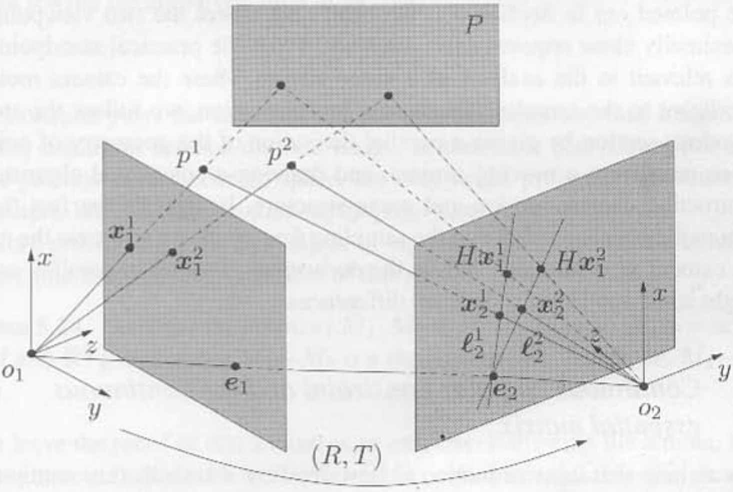

Figure 5.14. A homography H transfers two points $\boldsymbol{x}_1^1$ and $\boldsymbol{x}_1^2$ in the first image to two points $H\boldsymbol{x}_1^1$ and $H\boldsymbol{x}_1^2$ on the same epipolar lines as the respective true images $\boldsymbol{x}_2^1$ and $\boldsymbol{x}_2^2$ if the corresponding 3-D points p^1 and p^2 are not on the plane P from which H is induced.

Now consider the opposite situation that an essential matrix E is given and we want to compute the homography for a set of coplanar points. Note that once E is known, the vector T is also known (up to a scalar factor) as the left null space of E. We may typically choose T to be of unit length.

[9] Although in principle, to compute E from H, one does not need any extra information but only has to decompose H and find R and T using Theorem 5.19, the corollary will allow us to bypass that by much simpler techniques, which, unlike Theorem 5.19, will also be applicable to the uncalibrated case.

Corollary 5.23 (From essential matrix to homography). *Given an essential matrix E and three pairs of images $(\boldsymbol{x}_1^i, \boldsymbol{x}_2^i), i = 1, 2, 3,$ of three points in 3-D, the homography H induced by the plane specified by the three points then is*

$$H = \widehat{T}^T E + T v^T, \tag{5.47}$$

where $v = [v_1, v_2, v_3]^T \in \mathbb{R}^3$ solves the system of three linear equations

$$\widehat{\boldsymbol{x}_2^i}(\widehat{T}^T E + T v^T)\boldsymbol{x}_1^i = 0, \quad i = 1, 2, 3. \tag{5.48}$$

Proof. We leave the proof to the reader as an exercise. □

5.4 Continuous motion case[10]

As we pointed out in Section 5.1, the limit case where the two viewpoints are infinitesimally close requires extra attention. From the practical standpoint, this case is relevant to the analysis of a video stream where the camera motion is slow relative to the sampling frequency. In this section, we follow the steps of the previous section by giving a parallel derivation of the geometry of points in space as seen from a moving camera, and deriving a conceptual algorithm for reconstructing camera motion and scene structure. In light of the fact that the camera motion is slow relative to the sampling frequency, we will treat the motion of the camera as continuous. While the derivations proceed in parallel, we will highlight some subtle but significant differences.

5.4.1 Continuous epipolar constraint and the continuous essential matrix

Let us assume that camera motion is described by a smooth (i.e. continuously differentiable) trajectory $g(t) = (R(t), T(t)) \in SE(3)$ with body velocities $(\omega(t), v(t)) \in se(3)$ as defined in Chapter 2. For a point $p \in \mathbb{R}^3$, its coordinates as a function of time $\boldsymbol{X}(t)$ satisfy

$$\dot{\boldsymbol{X}}(t) = \widehat{\omega}(t)\boldsymbol{X}(t) + v(t). \tag{5.49}$$

The image of the point p taken by the camera is the vector $\boldsymbol{x}$ that satisfies $\lambda(t)\boldsymbol{x}(t) = \boldsymbol{X}(t)$. From now on, for convenience, we will drop the time dependency from the notation. Denote the velocity of the image point $\boldsymbol{x}$ by $\boldsymbol{u} \doteq \dot{\boldsymbol{x}} \in \mathbb{R}^3$. The velocity $\boldsymbol{u}$ is also called *image motion field*, which under the brightness constancy assumption discussed in Chapter 4 can be approximated by the *optical*

[10] This section can be skipped without loss of continuity if the reader is not interested in the continuous-motion case.

flow. To obtain an explicit expression for $\boldsymbol{u}$, we notice that

$$\boldsymbol{X} = \lambda \boldsymbol{x}, \quad \dot{\boldsymbol{X}} = \dot{\lambda}\boldsymbol{x} + \lambda\dot{\boldsymbol{x}}.$$

Substituting this into equation (5.49), we obtain

$$\dot{\boldsymbol{x}} = \widehat{\omega}\boldsymbol{x} + \frac{1}{\lambda}v - \frac{\dot{\lambda}}{\lambda}\boldsymbol{x}. \tag{5.50}$$

Then the image velocity $\boldsymbol{u} = \dot{\boldsymbol{x}}$ depends not only on the camera motion but also on the depth scale λ of the point. For the planar perspective projection and the spherical perspective projection, the expression for $\boldsymbol{u}$ will be slightly different. We leave the detail to the reader as an exercise (see Exercise 5.20).

To eliminate the depth scale λ, consider now the inner product of the vectors in (5.50) with the vector $(v \times \boldsymbol{x})$. We obtain

$$\dot{\boldsymbol{x}}^T\widehat{v}\boldsymbol{x} = \boldsymbol{x}^T\widehat{\omega}^T\widehat{v}\boldsymbol{x}.$$

We can rewrite the above equation in an equivalent way:

$$\boxed{\boldsymbol{u}^T\widehat{v}\boldsymbol{x} + \boldsymbol{x}^T\widehat{\omega}\widehat{v}\boldsymbol{x} = 0.} \tag{5.51}$$

This constraint plays the same role for the case of continuous-time images as the epipolar constraint for two discrete image, in the sense that it does not depend on the position of the point in space, but only on its projection and the motion parameters. We call it the *continuous epipolar constraint*.

Before proceeding with an analysis of equation (5.51), we state a lemma that will become useful in the remainder of this section.

Lemma 5.24. *Consider the matrices $M_1, M_2 \in \mathbb{R}^{3\times 3}$. Then $\boldsymbol{x}^TM_1\boldsymbol{x} = \boldsymbol{x}^TM_2\boldsymbol{x}$ for all $\boldsymbol{x} \in \mathbb{R}^3$ if and only if $M_1 - M_2$ is a skew-symmetric matrix, i.e. $M_1 - M_2 \in so(3)$.*

We leave the proof of this lemma as an exercise. Following the lemma, for any skew-symmetric matrix $M \in \mathbb{R}^{3\times 3}$, $\boldsymbol{x}^TM\boldsymbol{x} = 0$. Since $\frac{1}{2}(\widehat{\omega}\widehat{v} - \widehat{v}\widehat{\omega})$ is a skew-symmetric matrix, $\boldsymbol{x}^T\frac{1}{2}(\widehat{\omega}\widehat{v} - \widehat{v}\widehat{\omega})\boldsymbol{x} = 0$. If we define the *symmetric epipolar component* to be the matrix

$$s \doteq \frac{1}{2}(\widehat{\omega}\widehat{v} + \widehat{v}\widehat{\omega}) \quad \in \mathbb{R}^{3\times 3},$$

then we have that

$$\boldsymbol{x}^Ts\boldsymbol{x} = \boldsymbol{x}^T\widehat{\omega}\widehat{v}\boldsymbol{x},$$

so that the continuous epipolar constraint may be rewritten as

$$\boldsymbol{u}^T\widehat{v}\boldsymbol{x} + \boldsymbol{x}^Ts\boldsymbol{x} = 0. \tag{5.52}$$

This equation shows that for the matrix $\widehat{\omega}\widehat{v}$, only its symmetric component $s = \frac{1}{2}(\widehat{\omega}\widehat{v} + \widehat{v}\widehat{\omega})$ can be recovered from the epipolar equation (5.51) or equivalently

(5.52).[11] This structure is substantially different from that of the discrete case, and it cannot be derived by a first-order approximation of the essential matrix $\widehat{T}R$. In fact, a naive discretization of the discrete epipolar equation may lead to a constraint involving only a matrix of the form $\widehat{v}\widehat{\omega}$, whereas in the true continuous case we have to deal with only its symmetric component $s = \frac{1}{2}(\widehat{\omega}\widehat{v} + \widehat{v}\widehat{\omega})$ plus another term as given in (5.52). The set of matrices of interest in the case of continuous motions is thus the space of 6×3 matrices of the form

$$\mathcal{E}' \doteq \left\{ \begin{bmatrix} \widehat{v} \\ \frac{1}{2}(\widehat{\omega}\widehat{v} + \widehat{v}\widehat{\omega}) \end{bmatrix} \middle| \, \omega, v \in \mathbb{R}^3 \right\} \subset \mathbb{R}^{6\times 3},$$

which we call the *continuous essential space*. A matrix in this space is called a *continuous essential matrix*. Note that the continuous epipolar constraint (5.52) is homogeneous in the linear velocity v. Thus v may be recovered only up to a constant scalar factor. Consequently, in motion recovery, we will concern ourselves with matrices belonging to the *normalized continuous essential space* with v scaled to unit norm:

$$\mathcal{E}_1' = \left\{ \begin{bmatrix} \widehat{v} \\ \frac{1}{2}(\widehat{\omega}\widehat{v} + \widehat{v}\widehat{\omega}) \end{bmatrix} \middle| \, \omega \in \mathbb{R}^3, v \in \mathbb{S}^2 \right\} \subset \mathbb{R}^{6\times 3}.$$

5.4.2 *Properties of the continuous essential matrix*

The skew-symmetric part of a continuous essential matrix simply corresponds to the velocity v. The characterization of the (normalized) essential matrix focuses only on the symmetric matrix part $s = \frac{1}{2}(\widehat{\omega}\widehat{v} + \widehat{v}\widehat{\omega})$. We call the space of all the matrices of this form the *symmetric epipolar space*

$$\mathcal{S} \doteq \left\{ \frac{1}{2}(\widehat{\omega}\widehat{v} + \widehat{v}\widehat{\omega}) \,\middle|\, \omega \in \mathbb{R}^3, v \in \mathbb{S}^2 \right\} \subset \mathbb{R}^{3\times 3}.$$

The motion estimation problem is now reduced to that of *recovering the velocity (ω, v) with $\omega \in \mathbb{R}^3$ and $v \in \mathbb{S}^2$ from a given symmetric epipolar component s.*

The characterization of symmetric epipolar components depends on a characterization of matrices of the form $\widehat{\omega}\widehat{v} \in \mathbb{R}^{3\times 3}$, which is given in the following lemma. Of use in the lemma is the matrix $R_Y(\theta)$ defined to be the rotation around the Y-axis by an angle $\theta \in \mathbb{R}$, i.e. $R_Y(\theta) = e^{\widehat{e}_2\theta}$ with $e_2 = [0,1,0]^T \in \mathbb{R}^3$.

Lemma 5.25. *A matrix $Q \in \mathbb{R}^{3\times 3}$ has the form $Q = \widehat{\omega}\widehat{v}$ with $\omega \in \mathbb{R}^3$, $v \in \mathbb{S}^2$ if and only if*

$$Q = -VR_Y(\theta)diag\{\lambda, \lambda\cos(\theta), 0\}V^T \tag{5.53}$$

[11] This redundancy is the reason why different forms of the continuous epipolar constraint exist in the literature [Zhuang and Haralick, 1984, Ponce and Genc, 1998, Vieville and Faugeras, 1995, Maybank, 1993, Brooks et al., 1997], and accordingly, various approaches have been proposed to recover ω and v (see [Tian et al., 1996]).

for some rotation matrix $V \in SO(3)$*, the positive scalar* $\lambda = \|\omega\|$*, and* $\cos(\theta) = \omega^T v/\lambda$*.*

Proof. We first prove the necessity. The proof follows from the geometric meaning of $\widehat{\omega}\widehat{v}$ multiplied by any vector $q \in \mathbb{R}^3$:

$$\widehat{\omega}\widehat{v}q = \omega \times (v \times q).$$

Let $b \in \mathbb{S}^2$ be the unit vector perpendicular to both ω and v. That is, $b = \frac{v \times \omega}{\|v \times \omega\|}$. (If $v \times \omega = 0$, b is not uniquely defined. In this case, ω, v are parallel, and the rest of the proof follows if one picks any vector b orthogonal to v and ω.) Then $\omega = \lambda \exp(\widehat{b}\theta)v$ (according to this definition, θ is the angle between ω and v, and $0 \leq \theta \leq \pi$). It is easy to check that if the matrix V is defined to be

$$V = \left(e^{\widehat{b}\frac{\pi}{2}}v, b, v\right),$$

then Q has the given form (5.53).

We now prove the sufficiency. Given a matrix Q that can be decomposed into the form (5.53), define the orthogonal matrix $U = -VR_Y(\theta) \in O(3)$. (Recall that $O(3)$ represents the space of all orthogonal matrices of determinant ± 1.) Let the two skew-symmetric matrices $\widehat{\omega}$ and $\widehat{v}$ be given by

$$\widehat{\omega} = UR_Z\left(\pm\frac{\pi}{2}\right)\Sigma_\lambda U^T, \quad \widehat{v} = VR_Z\left(\pm\frac{\pi}{2}\right)\Sigma_1 V^T, \tag{5.54}$$

where $\Sigma_\lambda = \mathrm{diag}\{\lambda, \lambda, 0\}$ and $\Sigma_1 = \mathrm{diag}\{1, 1, 0\}$. Then

$$\begin{aligned} \widehat{\omega}\widehat{v} &= UR_Z\left(\pm\frac{\pi}{2}\right)\Sigma_\lambda U^T VR_Z\left(\pm\frac{\pi}{2}\right)\Sigma_1 V^T \\ &= UR_Z\left(\pm\frac{\pi}{2}\right)\Sigma_\lambda(-R_Y^T(\theta))R_Z\left(\pm\frac{\pi}{2}\right)\Sigma_1 V^T \\ &= U\mathrm{diag}\{\lambda, \lambda\cos(\theta), 0\}V^T \\ &= Q. \end{aligned} \tag{5.55}$$

Since ω and v have to be, respectively, the left and the right zero eigenvectors of Q, the reconstruction given in (5.54) is unique up to a sign. □

Based on the above lemma, the following theorem reveals the structure of the symmetric epipolar component.

Theorem 5.26 (Characterization of the symmetric epipolar component). *A real symmetric matrix* $s \in \mathbb{R}^{3\times 3}$ *is a symmetric epipolar component if and only if* s *can be diagonalized as* $s = V\Sigma V^T$ *with* $V \in SO(3)$ *and*

$$\Sigma = \mathit{diag}\{\sigma_1, \sigma_2, \sigma_3\}$$

with $\sigma_1 \geq 0, \sigma_3 \leq 0$*, and* $\sigma_2 = \sigma_1 + \sigma_3$*.*

Proof. We first prove the necessity. Suppose s is a symmetric epipolar component. Thus there exist $\omega \in \mathbb{R}^3, v \in \mathbb{S}^2$ such that $s = \frac{1}{2}(\widehat{\omega}\widehat{v} + \widehat{v}\widehat{\omega})$. Since s is a symmetric matrix, it is diagonalizable, all its eigenvalues are real, and all the eigenvectors are orthogonal to each other. It then suffices to check that its eigenvalues satisfy the given conditions.

Let the unit vector b, the rotation matrix V, θ, and λ be the same as in the proof of Lemma 5.25. According to the lemma, we have

$$\widehat{\omega}\widehat{v} = -VR_Y(\theta)\operatorname{diag}\{\lambda, \lambda\cos(\theta), 0\}V^T.$$

Since $(\widehat{\omega}\widehat{v})^T = \widehat{v}\widehat{\omega}$, we have

$$s = \frac{1}{2}V\left(-R_Y(\theta)\operatorname{diag}\{\lambda, \lambda\cos(\theta), 0\} - \operatorname{diag}\{\lambda, \lambda\cos(\theta), 0\}R_Y^T(\theta)\right)V^T.$$

Define the matrix $D(\lambda, \theta) \in \mathbb{R}^{3\times 3}$ to be

$$\begin{aligned} D(\lambda,\theta) &= -R_Y(\theta)\operatorname{diag}\{\lambda, \lambda\cos(\theta), 0\} - \operatorname{diag}\{\lambda, \lambda\cos(\theta), 0\}R_Y^T(\theta) \\ &= \lambda\begin{bmatrix} -2\cos(\theta) & 0 & \sin(\theta) \\ 0 & -2\cos(\theta) & 0 \\ \sin(\theta) & 0 & 0 \end{bmatrix}. \end{aligned}$$

Directly calculating its eigenvalues and eigenvectors, we obtain that $D(\lambda, \theta)$ is equal to

$$R_Y\left(\frac{\theta-\pi}{2}\right)\operatorname{diag}\{\lambda(1-\cos(\theta)), -2\lambda\cos(\theta), \lambda(-1-\cos(\theta))\}R_Y^T\left(\frac{\theta-\pi}{2}\right). \tag{5.56}$$

Thus $s = \frac{1}{2}VD(\lambda,\theta)V^T$ has eigenvalues

$$\left\{\frac{1}{2}\lambda(1-\cos(\theta)), \quad -\lambda\cos(\theta), \quad \frac{1}{2}\lambda(-1-\cos(\theta))\right\}, \tag{5.57}$$

which satisfy the given conditions.

We now prove the sufficiency. Given $s = V_1\operatorname{diag}\{\sigma_1, \sigma_2, \sigma_3\}V_1^T$ with $\sigma_1 \geq 0, \sigma_3 \leq 0, \sigma_2 = \sigma_1 + \sigma_3$, and $V_1^T \in SO(3)$, these three eigenvalues uniquely determine $\lambda, \theta \in \mathbb{R}$ such that the σ_i's have the form given in (5.57):

$$\begin{aligned} \lambda &= \sigma_1 - \sigma_3, & \lambda \geq 0, \\ \theta &= \arccos(-\sigma_2/\lambda), & \theta \in [0, \pi]. \end{aligned}$$

Define a matrix $V \in SO(3)$ to be $V = V_1R_Y^T\left(\frac{\theta}{2} - \frac{\pi}{2}\right)$. Then $s = \frac{1}{2}VD(\lambda,\theta)V^T$. According to Lemma 5.25, there exist vectors $v \in \mathbb{S}^2$ and $\omega \in \mathbb{R}^3$ such that

$$\widehat{\omega}\widehat{v} = -VR_Y(\theta)\operatorname{diag}\{\lambda, \lambda\cos(\theta), 0\}V^T.$$

Therefore, $\frac{1}{2}(\widehat{\omega}\widehat{v} + \widehat{v}\widehat{\omega}) = \frac{1}{2}VD(\lambda,\theta)V^T = s$. □

Figure 5.15 gives a geometric interpretation of the three eigenvectors of the symmetric epipolar component s for the case in which both ω, v are of unit length. The constructive proof given above is important since it gives an explicit decomposition of the symmetric epipolar component s, which will be studied in more detail next.

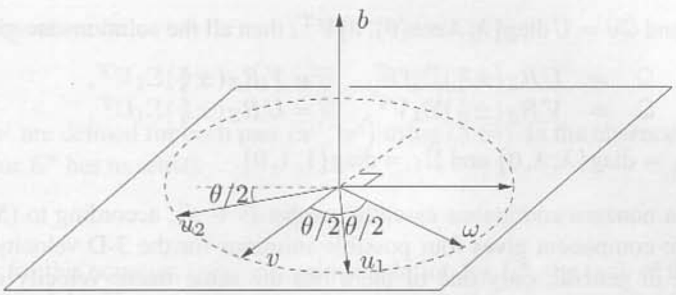

Figure 5.15. Vectors u_1, u_2, b are the three eigenvectors of a symmetric epipolar component $\frac{1}{2}(\widehat{\omega}\widehat{v}+\widehat{v}\widehat{\omega})$. In particular, b is the normal vector to the plane spanned by ω and v, and u_1, u_2 are both in this plane. The vector u_1 is the average of ω and v, and u_2 is orthogonal to both b and u_1.

Following the proof of Theorem 5.26, if we already know the eigenvector decomposition of a symmetric epipolar component s, we certainly can find at least one solution (ω, v) such that $s = \frac{1}{2}(\widehat{\omega}\widehat{v} + \widehat{v}\widehat{\omega})$. We now discuss uniqueness, i.e. how many solutions exist for $s = \frac{1}{2}(\widehat{\omega}\widehat{v} + \widehat{v}\widehat{\omega})$.

Theorem 5.27 (Velocity recovery from the symmetric epipolar component). *There exist exactly four 3-D velocities (ω, v) with $\omega \in \mathbb{R}^3$ and $v \in \mathbb{S}^2$ corresponding to a nonzero $s \in \mathcal{S}$.*

Proof. Suppose (ω_1, v_1) and (ω_2, v_2) are both solutions for $s = \frac{1}{2}(\widehat{\omega}\widehat{v} + \widehat{v}\widehat{\omega})$. Then we have

$$\widehat{v}_1\widehat{\omega}_1 + \widehat{\omega}_1\widehat{v}_1 = \widehat{v}_2\widehat{\omega}_2 + \widehat{\omega}_2\widehat{v}_2. \tag{5.58}$$

From Lemma 5.25, we may write

$$\begin{aligned} \widehat{\omega}_1\widehat{v}_1 &= -V_1 R_Y(\theta_1)\mathrm{diag}\{\lambda_1, \lambda_1\cos(\theta_1), 0\}V_1^T, \\ \widehat{\omega}_2\widehat{v}_2 &= -V_2 R_Y(\theta_2)\mathrm{diag}\{\lambda_2, \lambda_2\cos(\theta_2), 0\}V_2^T. \end{aligned} \tag{5.59}$$

Let $W = V_1^T V_2 \in SO(3)$. Then from (5.58),

$$D(\lambda_1, \theta_1) = W D(\lambda_2, \theta_2) W^T. \tag{5.60}$$

Since both sides of (5.60) have the same eigenvalues, according to (5.56), we have

$$\lambda_1 = \lambda_2, \quad \theta_2 = \theta_1.$$

We can then denote both θ_1 and θ_2 by θ. It is immediate to check that the only possible rotation matrix W that satisfies (5.60) is given by $I_{3\times 3}$,

$$\begin{bmatrix} -\cos(\theta) & 0 & \sin(\theta) \\ 0 & -1 & 0 \\ \sin(\theta) & 0 & \cos(\theta) \end{bmatrix}, \quad \text{or} \quad \begin{bmatrix} \cos(\theta) & 0 & -\sin(\theta) \\ 0 & -1 & 0 \\ -\sin(\theta) & 0 & -\cos(\theta) \end{bmatrix}.$$

From the geometric meaning of V_1 and V_2, all the cases give either $\widehat{\omega}_1\widehat{v}_1 = \widehat{\omega}_2\widehat{v}_2$ or $\widehat{\omega}_1\widehat{v}_1 = \widehat{v}_2\widehat{\omega}_2$. Thus, according to the proof of Lemma 5.25, if (ω, v) is one solution and $\widehat{\omega}\widehat{v} = U\text{diag}\{\lambda, \lambda\cos(\theta), 0\}V^T$, then all the solutions are given by

$$\begin{array}{rcl} \widehat{\omega} & = & UR_Z(\pm\frac{\pi}{2})\Sigma_\lambda U^T, \quad \widehat{v} = VR_Z(\pm\frac{\pi}{2})\Sigma_1 V^T, \\ \widehat{\omega} & = & VR_Z(\pm\frac{\pi}{2})\Sigma_\lambda V^T, \quad \widehat{v} = UR_Z(\pm\frac{\pi}{2})\Sigma_1 U^T, \end{array} \tag{5.61}$$

where $\Sigma_\lambda = \text{diag}\{\lambda, \lambda, 0\}$ and $\Sigma_1 = \text{diag}\{1, 1, 0\}$. □

Given a nonzero continuous essential matrix $E \in \mathcal{E}'$, according to (5.61), its symmetric component gives four possible solutions for the 3-D velocity (ω, v). However, in general, only one of them has the same linear velocity v as the skew-symmetric part of E. Hence, compared to the discrete case, where there are two 3-D motions (R, T) associated with an essential matrix, the velocity (ω, v) corresponding to a continuous essential matrix is unique. This is because in the continuous case, the *twisted-pair ambiguity*, which occurs in the discrete case and is caused by a 180° rotation of the camera around the translation direction, see Example 5.8, is now avoided.

5.4.3 The eight-point linear algorithm

Based on the preceding study of the continuous essential matrix, this section describes an algorithm to recover the 3-D velocity of the camera from a set of (possibly noisy) optical flow measurements.

Let $E = \begin{bmatrix} \widehat{v} \\ s \end{bmatrix} \in \mathcal{E}_1'$ with $s = \frac{1}{2}(\widehat{\omega}\widehat{v} + \widehat{v}\widehat{\omega})$ be the essential matrix associated with the continuous epipolar constraint (5.52). Since the submatrix $\widehat{v}$ is skew-symmetric and s is symmetric, they have the following form

$$\widehat{v} = \begin{bmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{bmatrix}, \quad s = \begin{bmatrix} s_1 & s_2 & s_3 \\ s_2 & s_4 & s_5 \\ s_3 & s_5 & s_6 \end{bmatrix}. \tag{5.62}$$

Define the continuous version of the "stacked" vector $E^s \in \mathbb{R}^9$ to be

$$E^s \doteq [v_1, v_2, v_3, s_1, s_2, s_3, s_4, s_5, s_6]^T. \tag{5.63}$$

Define a vector $\boldsymbol{a} \in \mathbb{R}^9$ associated with the optical flow $(\boldsymbol{x}, \boldsymbol{u})$ with $\boldsymbol{x} = [x, y, z]^T \in \mathbb{R}^3$, $\boldsymbol{u} = [u_1, u_2, u_3]^T \in \mathbb{R}^3$ to be[15]

$$\boldsymbol{a} \doteq [u_3y - u_2z, u_1z - u_3x, u_2x - u_1y, x^2, 2xy, 2xz, y^2, 2yz, z^2]^T. \tag{5.64}$$

The continuous epipolar constraint (5.52) can be then rewritten as

$$\boldsymbol{a}^T E^s = 0.$$

Given a set of (possibly noisy) optical flow vectors $(\boldsymbol{x}^j, \boldsymbol{u}^j)$, $j = 1, 2, \ldots, n$, generated by the same motion, define a matrix $\chi \in \mathbb{R}^{n \times 9}$ associated with these

[15] For a planar perspective projection, $z = 1$ and $u_3 = 0$; thus the expression for $\boldsymbol{a}$ can be simplified.

measurements to be

$$\chi \doteq [a^1, a^2, \ldots, a^n]^T, \tag{5.65}$$

where a^j are defined for each pair (x^j, u^j) using (5.64). In the absence of noise, the vector E^s has to satisfy

$$\chi E^s = 0. \tag{5.66}$$

In order for this equation to have a unique solution for E^s, the rank of the matrix χ has to be eight. Thus, *for this algorithm, the optical flow vectors of at least eight points are needed to recover the 3-D velocity, i.e.* $n \geq 8$, although the minimum number of optical flow vectors needed for a finite number of solutions is actually five, as discussed by [Maybank, 1993].

When the measurements are noisy, there may be no solution to $\chi E^s = 0$. As in the discrete case, one may approximate the solution by minimizing the least-squares error function $\|\chi E^s\|^2$.

Since the vector E^s is recovered from noisy measurements, the symmetric part s of E directly recovered from unstacking E^s is not necessarily a symmetric epipolar component. Thus one cannot directly use the previously derived results for symmetric epipolar components to recover the 3-D velocity. In analogy to the discrete case, we can project the symmetric matrix s onto the space of symmetric epipolar components.

Theorem 5.28 (Projection onto the symmetric epipolar space). *If a real symmetric matrix $F \in \mathbb{R}^{3\times3}$ is diagonalized as $F = V\mathrm{diag}\{\lambda_1, \lambda_2, \lambda_3\}V^T$ with $V \in SO(3)$, $\lambda_1 \geq 0, \lambda_3 \leq 0$, and $\lambda_1 \geq \lambda_2 \geq \lambda_3$, then the symmetric epipolar component $E \in \mathcal{S}$ that minimizes the error $\|E - F\|_f^2$ is given by $E = V\mathrm{diag}\{\sigma_1, \sigma_2, \sigma_2\}V^T$ with*

$$\sigma_1 = \frac{2\lambda_1 + \lambda_2 - \lambda_3}{3}, \quad \sigma_2 = \frac{\lambda_1 + 2\lambda_2 + \lambda_3}{3}, \quad \sigma_3 = \frac{2\lambda_3 + \lambda_2 - \lambda_1}{3}. \tag{5.67}$$

Proof. Define $\mathcal{S}_\Sigma$ to be the subspace of $\mathcal{S}$ whose elements have the same eigenvalues $\Sigma = \mathrm{diag}\{\sigma_1, \sigma_2, \sigma_3\}$. Thus every matrix $E \in \mathcal{S}_\Sigma$ has the form $E = V_1\Sigma V_1^T$ for some $V_1 \in SO(3)$. To simplify the notation, define $\Sigma_\lambda = \mathrm{diag}\{\lambda_1, \lambda_2, \lambda_3\}$. We now prove this theorem in two steps.

Step 1: We prove that the matrix $E \in \mathcal{S}_\Sigma$ that minimizes the error $\|E - F\|_f^2$ is given by $E = V\Sigma V^T$. Since $E \in \mathcal{S}_\Sigma$ has the form $E = V_1\Sigma V_1^T$, we get

$$\|E - F\|_f^2 = \|V_1\Sigma V_1^T - V\Sigma_\lambda V^T\|_f^2 = \|\Sigma_\lambda - V^T V_1 \Sigma V_1^T V\|_f^2.$$

Define $W = V^T V_1 \in SO(3)$ and denote its entries by

$$W = \begin{bmatrix} w_1 & w_2 & w_3 \\ w_4 & w_5 & w_6 \\ w_7 & w_8 & w_9 \end{bmatrix}. \tag{5.68}$$

Then

$$\begin{aligned} \|E - F\|_f^2 &= \|\Sigma_\lambda - W\Sigma W^T\|_f^2 \\ &= \operatorname{trace}(\Sigma_\lambda^2) - 2\operatorname{trace}(W\Sigma W^T\Sigma_\lambda) + \operatorname{trace}(\Sigma^2). \end{aligned} \tag{5.69}$$

Substituting (5.68) into the second term, and using the fact that $\sigma_2 = \sigma_1 + \sigma_3$ and W is a rotation matrix, we get

$$\begin{aligned} \operatorname{trace}(W\Sigma W^T\Sigma_\lambda) &= \sigma_1(\lambda_1(1 - w_3^2) + \lambda_2(1 - w_6^2) + \lambda_3(1 - w_9^2)) \\ &\quad + \sigma_3(\lambda_1(1 - w_1^2) + \lambda_2(1 - w_4^2) + \lambda_3(1 - w_7^2)). \end{aligned}$$

Minimizing $\|E - F\|_f^2$ is equivalent to maximizing $\operatorname{trace}(W\Sigma W^T\Sigma_\lambda)$. From the above equation, $\operatorname{trace}(W\Sigma W^T\Sigma_\lambda)$ is maximized if and only if $w_3 = w_6 = 0$, $w_9^2 = 1$, $w_4 = w_7 = 0$, and $w_1^2 = 1$. Since W is a rotation matrix, we also have $w_2 = w_8 = 0$, and $w_5^2 = 1$. All possible W give a unique matrix in $\mathcal{S}_\Sigma$ that minimizes $\|E - F\|_f^2$: $E = V\Sigma V^T$.

Step 2: From step one, we need only to minimize the error function over the matrices that have the form $V\Sigma V^T \in \mathcal{S}$. The optimization problem is then converted to one of minimizing the error function

$$\|E - F\|_f^2 = (\lambda_1 - \sigma_1)^2 + (\lambda_2 - \sigma_2)^2 + (\lambda_3 - \sigma_3)^2$$

subject to the constraint

$$\sigma_2 = \sigma_1 + \sigma_3.$$

The formulae (5.67) for $\sigma_1, \sigma_2, \sigma_3$ are directly obtained from solving this minimization problem. □

Remark 5.29. *In the preceding theorem, for a symmetric matrix F that does not satisfy the conditions $\lambda_1 \geq 0$ and $\lambda_3 \leq 0$, one chooses $\lambda_1' = \max\{\lambda_1, 0\}$ and $\lambda_3' = \min\{\lambda_3, 0\}$ prior to applying the above theorem.*

Finally, we outline an eigenvalue-decomposition algorithm, Algorithm 5.3, for estimating 3-D velocity from optical flows of eight points, which serves as a continuous counterpart of the eight-point algorithm given in Section 5.2.

Remark 5.30. *Since both $E, -E \in \mathcal{E}_1'$ satisfy the same set of continuous epipolar constraints, both $(\omega, \pm v)$ are possible solutions for the given set of optical flow vectors. However, as in the discrete case, one can get rid of the ambiguous solution by enforcing the positive depth constraint (Exercise 5.11).*

In situations where the motion of the camera is partially constrained, the above linear algorithm can be further simplified. The following example illustrates such a scenario.

Example 5.31 (Constrained motion estimation). This example shows how to utilize constraints on motion to be estimated in order to simplify the proposed linear motion estimation algorithm in the continuous case. Let $g(t) \in SE(3)$ represent the position and orientation of an aircraft relative to the spatial frame; the inputs $\omega_1, \omega_2, \omega_3 \in \mathbb{R}$ stand for

Algorithm 5.3 (The continuous eight-point algorithm).

For a given set of images and optical flow vectors (x^j, u^j), $j = 1, 2, \ldots, n$, this algorithm finds $(\omega, v) \in SE(3)$ that solves

$$u^{jT}\widehat{v}x^j + x^{jT}\widehat{\omega}\widehat{v}x^j = 0, \quad j = 1, 2, \ldots, n.$$

1. **Estimate the essential vector**
   Define a matrix $\chi \in \mathbb{R}^{n\times 9}$ whose jth row is constructed from x^j and u^j as in (5.64). Use the SVD to find the vector $E^s \in \mathbb{R}^9$ such that $\chi E^s = 0$: $\chi = U_\chi \Sigma_\chi V_\chi^T$ and $E^s = V_\chi(:, 9)$. Recover the vector $v_0 \in \mathbb{S}^2$ from the first three entries of E^s and a symmetric matrix $s \in \mathbb{R}^{3\times 3}$ from the remaining six entries as in (5.63). Multiply E^s with a scalar such that the vector v_0 becomes of unit norm.

2. **Recover the symmetric epipolar component**
   Find the eigenvalue decomposition of the symmetric matrix s:

   $$s = V_1 \mathrm{diag}\{\lambda_1, \lambda_2, \lambda_3\} V_1^T,$$

   with $\lambda_1 \geq \lambda_2 \geq \lambda_3$. Project the symmetric matrix s onto the symmetric epipolar space $\mathcal{S}$. We then have the new $s = V_1 \mathrm{diag}\{\sigma_1, \sigma_2, \sigma_3\} V_1^T$ with

   $$\sigma_1 = \frac{2\lambda_1 + \lambda_2 - \lambda_3}{3}, \; \sigma_2 = \frac{\lambda_1 + 2\lambda_2 + \lambda_3}{3}, \; \sigma_3 = \frac{2\lambda_3 + \lambda_2 - \lambda_1}{3}.$$

3. **Recover the velocity from the symmetric epipolar component**
   Define

   $$\begin{aligned} \lambda &= \sigma_1 - \sigma_3, \quad \lambda \geq 0, \\ \theta &= \arccos(-\sigma_2/\lambda), \quad \theta \in [0, \pi]. \end{aligned}$$

   Let $V = V_1 R_Y^T\left(\frac{\theta}{2} - \frac{\pi}{2}\right) \in SO(3)$ and $U = -VR_Y(\theta) \in O(3)$. Then the four possible 3-D velocities corresponding to the matrix s are given by

   $$\begin{aligned} \widehat{\omega} &= UR_Z(\pm\tfrac{\pi}{2})\Sigma_\lambda U^T, \quad \widehat{v} = VR_Z(\pm\tfrac{\pi}{2})\Sigma_1 V^T, \\ \widehat{\omega} &= VR_Z(\pm\tfrac{\pi}{2})\Sigma_\lambda V^T, \quad \widehat{v} = UR_Z(\pm\tfrac{\pi}{2})\Sigma_1 U^T, \end{aligned}$$

   where $\Sigma_\lambda = \mathrm{diag}\{\lambda, \lambda, 0\}$ and $\Sigma_1 = \mathrm{diag}\{1, 1, 0\}$.

4. **Recover velocity from the continuous essential matrix**
   From the four velocities recovered from the matrix s in step 3, choose the pair (ω^*, v^*) that satisfies

   $$v^{*T}v_0 = \max_i\{v_i^T v_0\}.$$

   Then the estimated 3-D velocity (ω, v) with $\omega \in \mathbb{R}^3$ and $v \in \mathbb{S}^2$ is given by

   $$\omega = \omega^*, \quad v = v_0.$$

the rates of the rotation about the axes of the aircraft, and $v_1 \in \mathbb{R}$ is the velocity of the aircraft. Using the standard homogeneous representation for g (see Chapter 2), the kinematic equations of the aircraft motion are given by

$$\dot{g} \;=\; \begin{bmatrix} 0 & -\omega_3 & \omega_2 & v_1 \\ \omega_3 & 0 & -\omega_1 & 0 \\ -\omega_2 & \omega_1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} g,$$

where ω_1 stands for pitch rate, ω_2 for roll rate, ω_3 for yaw rate, and v_1 for the velocity of the aircraft. Then the 3-D velocity (ω, v) in the continuous epipolar constraint (5.52) has the form $\omega = [\omega_1, \omega_2, \omega_3]^T, v = [v_1, 0, 0]^T$. For Algorithm 5.3, we have extra constraints on the symmetric matrix $s = \frac{1}{2}(\widehat{\omega}\widehat{v} + \widehat{v}\widehat{\omega})$: $s_1 = s_5 = 0$ and $s_4 = s_6$. Then there are only four different essential parameters left to determine, and we can redefine the motion parameter vector $E^s \in \mathbb{R}^4$ to be $E^s \doteq [v_1, s_2, s_3, s_4]^T$. Then the measurement vector $\boldsymbol{a} \in \mathbb{R}^4$ is given by $\boldsymbol{a} = [u_3 y - u_2 z, 2xy, 2xz, y^2 + z^2]^T$. The continuous epipolar constraint can then be rewritten as

$$\boldsymbol{a}^T E^s = 0.$$

If we define the matrix χ from $\boldsymbol{a}$ as in (5.65), the matrix $\chi^T \chi$ is a 4×4 matrix rather than a 9×9. For estimating the velocity $(\omega; v)$, the dimension of the problem is then reduced from nine to four. In this special case, the minimum number of optical flow measurements needed to guarantee a unique solution of E^s is reduced to four instead of eight. Furthermore, the symmetric matrix s recovered from E^s is automatically in the space $\mathcal{S}$, and the remaining steps of the algorithm can thus be dramatically simplified. From this simplified algorithm, the angular velocity $\omega = [\omega_1, \omega_2, \omega_3]^T$ can be fully recovered from the images. The velocity information can then be used for controlling the aircraft. ■

As in the discrete case, the linear algorithm proposed above is not optimal, since it does not enforce the structure of the parameter space during the minimization. Therefore, the recovered velocity does not necessarily minimize the originally chosen error function $\|\chi E^s(\omega, v)\|^2$ on the space $\mathcal{E}_1'$.

Additionally, as in the discrete case, we have to assume that translation is not zero. If the motion is purely rotational, then one can prove that there are infinitely many solutions to the epipolar constraint-related equations. We leave this as an exercise to the reader.

5.4.4 Euclidean constraints and structure reconstruction

As in the discrete case, the purpose of exploiting Euclidean constraints is to reconstruct the scales of the motion and structure. From the above linear algorithm, we know that we can recover the linear velocity v only up to an arbitrary scalar factor. Without loss of generality, we may assume that the velocity of the camera motion to be $(\omega, \eta v)$ with $\|v\| = 1$ and $\eta \in \mathbb{R}$. By now, only the scale factor η is unknown. Substituting $\boldsymbol{X}(t) = \lambda(t)\boldsymbol{x}(t)$ into the equation

$$\dot{\boldsymbol{X}}(t) = \widehat{\omega}\boldsymbol{X}(t) + \eta v(t),$$

we obtain for the image $\boldsymbol{x}^j$ of each point $p^j \in \mathbb{E}^3, j = 1, 2, \ldots, n,$

$$\dot{\lambda}^j \boldsymbol{x}^j + \lambda^j \dot{\boldsymbol{x}}^j = \widehat{\omega}(\lambda^j \boldsymbol{x}^j) + \eta v \;\Leftrightarrow\; \dot{\lambda}^j \boldsymbol{x}^j + \lambda^j(\dot{\boldsymbol{x}}^j - \widehat{\omega}\boldsymbol{x}^j) - \eta v = 0. \quad (5.70)$$

As one may expect, in the continuous case, the scale information is then encoded in $\lambda, \dot{\lambda}$ for the location of the 3-D point, and $\eta \in \mathbb{R}^+$ for the linear velocity v. Knowing $\boldsymbol{x}, \dot{\boldsymbol{x}}, \omega$, and v, we see that these constraints are all linear in $\lambda^j, \dot{\lambda}^j, 1 \leq j \leq n$, and η. Also, if $\boldsymbol{x}^j, 1 \leq j \leq n$ are linearly independent of v, i.e. the feature points do not line up with the direction of translation, it can be shown that these linear constraints are not degenerate; hence the unknown scales are determined up to a universal scalar factor. We may then arrange all the unknown scalars into a single vector $\vec{\lambda}$:

$$\vec{\lambda} = [\lambda^1, \lambda^2 \ldots, \lambda^n, \dot{\lambda}^1, \dot{\lambda}^2, \ldots, \dot{\lambda}^n, \eta]^T \quad \in \mathbb{R}^{2n+1}.$$

For n optical flow vectors, $\vec{\lambda}$ is a $(2n+1)$-dimensional vector. (5.70) gives $3n$ (scalar) linear equations. The problem of solving $\vec{\lambda}$ from (5.70) is usually over determined. It is easy to check that in the absence of noise the set of equations given by (5.70) uniquely determines $\vec{\lambda}$ if the configuration is noncritical. We can therefore write all the equations in the matrix form

$$M\vec{\lambda} = 0,$$

with $M \in \mathbb{R}^{3n \times (2n+1)}$ a matrix depending on ω, v, and $\{(\boldsymbol{x}^j, \dot{\boldsymbol{x}}^j)\}_{j=1}^n$. Then, in the presence of noise, the linear least-squares estimate of $\vec{\lambda}$ is simply the eigenvector of $M^T M$ corresponding to the smallest eigenvalue.

Notice that the time derivative of the scales $\{\dot{\lambda}^j\}_{j=1}^n$ can also be estimated. Suppose we have done the above recovery for a time interval, say (t_0, t_f). Then we have the estimate $\vec{\lambda}(t)$ as a function of time t. But $\vec{\lambda}(t)$ at each time t is determined only up to an arbitrary scalar factor. Hence $\rho(t)\vec{\lambda}(t)$ is also a valid estimation for any positive function $\rho(t)$ defined on (t_0, t_f). However, since $\rho(t)$ is multiplied by both $\lambda(t)$ and $\dot{\lambda}(t)$, their ratio

$$r(t) = \dot{\lambda}(t)/\lambda(t)$$

is independent of the choice of $\rho(t)$. Notice that $\frac{d}{dt}(\ln \lambda) = \dot{\lambda}/\lambda$. Let the logarithm of the structural scale λ be $y = \ln \lambda$. Then a time-consistent estimation $\lambda(t)$ needs to satisfy the following ordinary differential equation, which we call the *dynamical scale ODE*

$$\dot{y}(t) = r(t).$$

Given $y(t_0) = y_0 = \ln(\lambda(t_0))$, we solve this ODE and obtain $y(t)$ for $t \in [t_0, t_f]$. Then we can recover a consistent scale $\lambda(t)$ given by

$$\lambda(t) = \exp(y(t)).$$

Hence (structure and motion) scales estimated at different time instances now are all relative to the same scale at time t_0. Therefore, in the continuous case, we are also able to recover all the scales as functions of time up to a universal scalar factor. The reader must be aware that the above scheme is only *conceptual*. In reality, the ratio function $r(t)$ would never be available for *every* time instant in $[t_0, t_f]$.

Universal scale ambiguity

In both the discrete and continuous cases, in principle, the proposed schemes can reconstruct both the Euclidean structure and motion up to a universal scalar factor. This ambiguity is intrinsic, since one can scale the entire world up or down with a scaling factor while all the images obtained remain the same. In all the algorithms proposed above, this factor is fixed (rather arbitrarily, in fact) by imposing the translation scale to be 1. In practice, this scale and its unit can also be chosen to be directly related to some known length, size, distance, or motion of an object in space.

5.4.5 Continuous homography for a planar scene

In this section, we consider the continuous version of the case that we have studied in Section 5.3, where all the feature points of interest are lying on a plane P. Planar scenes are a degenerate case for the discrete epipolar constraint, and also for the continuous case. Recall that in the continuous scenario, instead of having image pairs, we measure the image point $\boldsymbol{x}$ and its optical flow $\boldsymbol{u} = \dot{\boldsymbol{x}}$. Other assumptions are the same as in Section 5.3.

Suppose the camera undergoes a rigid-body motion with body angular and linear velocities ω, v. Then the time derivative of coordinates $\boldsymbol{X} \in \mathbb{R}^3$ of a point p (with respect to the camera frame) satisfies[19]

$$\dot{\boldsymbol{X}} = \widehat{\omega}\boldsymbol{X} + v. \tag{5.71}$$

Let $N \in \mathbb{R}^3$ be the surface normal to P (with respect to the camera frame) at time t. Then, if $d(t) > 0$ is the distance from the optical center of the camera to the plane P at time t, then

$$N^T\boldsymbol{X} = d \quad \Leftrightarrow \quad \frac{1}{d}N^T\boldsymbol{X} = 1, \quad \forall \boldsymbol{X} \in P. \tag{5.72}$$

Substituting equation (5.72) into equation (5.71) yields the relation

$$\dot{\boldsymbol{X}} = \widehat{\omega}\boldsymbol{X} + v = \widehat{\omega}\boldsymbol{X} + v\frac{1}{d}N^T\boldsymbol{X} = \left(\widehat{\omega} + \frac{1}{d}vN^T\right)\boldsymbol{X}. \tag{5.73}$$

As in the discrete case, we call the matrix

$$H \doteq \left(\widehat{\omega} + \frac{1}{d}vN^T\right) \in \mathbb{R}^{3\times 3} \tag{5.74}$$

the *continuous homography matrix*. For simplicity, here we use the same symbol H to denote it, and it really is a continuous (or infinitesimal) version of the (discrete) homography matrix $H = R + \frac{1}{d}TN^T$ studied in Section 5.3.

[19] Here, as in previous cases, we assume implicitly that time dependency of $\boldsymbol{X}$ on t is smooth so that we can take derivatives whenever necessary. However, for simplicity, we drop the dependency of $\boldsymbol{X}$ on t in the notation $\boldsymbol{X}(t)$.

Note that the matrix H depends both on the continuous motion parameters $\{\omega, v\}$ and structure parameters $\{N, d\}$ that we wish to recover. As in the discrete case, there is an inherent scale ambiguity in the term $\frac{1}{d}v$ in equation (5.74). Thus, in general, knowing H, one can recover only the ratio of the camera translational velocity scaled by the distance to the plane.

From the relation

$$\lambda \boldsymbol{x} = \boldsymbol{X}, \quad \dot{\lambda}\boldsymbol{x} + \lambda \boldsymbol{u} = \dot{\boldsymbol{X}}, \quad \dot{\boldsymbol{X}} = H\boldsymbol{X}, \tag{5.75}$$

we have

$$\boxed{\boldsymbol{u} = H\boldsymbol{x} - \frac{\dot{\lambda}}{\lambda}\boldsymbol{x}.} \tag{5.76}$$

This is indeed the continuous version of the planar homography.

5.4.6 Estimating the continuous homography matrix

In order to further eliminate the depth scale λ in equation (5.76), multiplying both sides by the skew-symmetric matrix $\widehat{\boldsymbol{x}} \in \mathbb{R}^{3\times 3}$, we obtain the equation

$$\boxed{\widehat{\boldsymbol{x}} H \boldsymbol{x} = \widehat{\boldsymbol{x}} \boldsymbol{u}.} \tag{5.77}$$

We may call this the *continuous homography constraint* or the *continuous planar epipolar constraint* as a continuous version of the discrete case.

Since this constraint is linear in H, by stacking the entries of H as

$$H^s = [H_{11}, H_{21}, H_{31}, H_{12}, H_{22}, H_{32}, H_{13}, H_{23}, H_{33}]^T \quad \in \mathbb{R}^9,$$

we may rewrite (5.77) as

$$\boldsymbol{a}^T H^s = \widehat{\boldsymbol{x}}\boldsymbol{u},$$

where $\boldsymbol{a} \in \mathbb{R}^{9\times 3}$ is the Kronecker product $\boldsymbol{x} \otimes \widehat{\boldsymbol{x}}$. However, since the skew-symmetric matrix $\widehat{\boldsymbol{x}}$ is only of rank 2, the equation imposes only two constraints on the entries of H. Given a set of n image point and velocity pairs $\{(\boldsymbol{x}^j, \boldsymbol{u}^j)\}_{j=1}^n$ of points on the plane, we may stack all equations $\boldsymbol{a}^{jT} H^s = \widehat{\boldsymbol{x}^j \boldsymbol{u}^j}, j = 1, 2, \ldots, n$, into a single equation

$$\chi H^s = B, \tag{5.78}$$

where $\chi \doteq \left[\boldsymbol{a}^1, \ldots, \boldsymbol{a}^n\right]^T \in \mathbb{R}^{3n\times 9}$ and $B \doteq \left[(\widehat{\boldsymbol{x}^1 \boldsymbol{u}^1})^T, \ldots, (\widehat{\boldsymbol{x}^j \boldsymbol{u}^j})^T\right]^T \in \mathbb{R}^{3n}$.

In order to solve uniquely (up to a scalar factor) for H^s, we must have $\text{rank}(\chi) = 8$. Since each pair of image points gives two constraints, we expect that at least four optical flow pairs would be necessary for a unique estimate of H (up to a scalar factor). In analogy with the discrete case, we have the following statement, the proof of which we leave to the reader as a linear-algebra exercise.

Proposition 5.32 (Four-point continuous homography). *We have $rank(\chi) = 8$ if and only if there exists a set of four points (out of the n) such that any three of them are not collinear; i.e. they are in general configuration in the plane.*

Then, if optical flow at more than four points in general configuration in the plane is given, using linear least-squares techniques, equation (5.78) can be used to recover H^s up to one dimension, since χ has a one-dimensional null space. That is, we can recover $H_L = H - \xi H_K$, where H_L corresponds to the minimum-norm linear least-squares estimate of H solving $\min \|\chi H^s - B\|^2$, and H_K corresponds to a vector in null(χ) and $\xi \in \mathbb{R}$ is an unknown scalar factor.

By inspection of equation (5.77) one can see that $H_K = I$, since $\widehat{x} I x = \widehat{x} x = 0$. Then we have

$$H = H_L + \xi I. \tag{5.79}$$

Thus, in order to recover H, we need only to identify the unknown ξ. So far, we have not considered the special structure of the matrix H. Next, we give constraints imposed by the structure of H that will allow us to identify ξ, and thus uniquely recover H.

Lemma 5.33. *Suppose $u, v \in \mathbb{R}^3$, and $\|u\|^2 = \|v\|^2 = \alpha$. If $u \neq v$, the matrix $D = uv^T + vu^T \in \mathbb{R}^{3\times 3}$ has eigenvalues $\{\lambda_1, 0, \lambda_3\}$, where $\lambda_1 > 0$, and $\lambda_3 < 0$. If $u = \pm v$, the matrix D has eigenvalues $\{\pm 2\alpha, 0, 0\}$.*

Proof. Let $\beta = u^T v$. If $u \neq \pm v$, we have $-\alpha < \beta < \alpha$. We can solve the eigenvalues and eigenvectors of D by

$$\begin{aligned} D(u+v) &= (\beta+\alpha)(u+v), \\ D(u\times v) &= 0, \\ D(u-v) &= (\beta-\alpha)(u-v). \end{aligned}$$

Clearly, $\lambda_1 = (\beta+\alpha) > 0$ and $\lambda_3 = \beta - \alpha < 0$. It is easy to check the conditions on D when $u = \pm v$. □

Lemma 5.34 (Normalization of the continuous homography matrix). *Given the H_L part of a continuous planar homography matrix of the form $H = H_L + \xi I$, we have*

$$\xi = -\frac{1}{2}\gamma_2(H_L + H_L^T), \tag{5.80}$$

where $\gamma_2(H_L + H_L^T) \in \mathbb{R}$ is the second-largest eigenvalue of $H_L + H_L^T$.

Proof. In this proof, we will work with sorted eigenvalues; that is, if $\{\lambda_1, \lambda_2, \lambda_3\}$ are eigenvalues of some matrix, then $\lambda_1 \geq \lambda_2 \geq \lambda_3$. If the points are not in general configuration, then rank$(\chi) < 7$, and the problem is under constrained. Now suppose the points are in general configuration. Then by least-squares estimation we may recover $H_L = H - \xi I$ for some unknown $\xi \in \mathbb{R}$. By Lemma 5.33, $H + H^T = \frac{1}{d} v N^T + \frac{1}{d} N v^T$ has eigenvalues $\{\lambda_1, \lambda_2, \lambda_3\}$, where $\lambda_1 \geq 0$, $\lambda_2 = 0$, and $\lambda_3 \leq 0$. So compute the eigenvalues of $H_L + H_L^T$ and denote them by $\{\gamma_1, \gamma_2, \gamma_3\}$. Since we have $H = H_L + \xi I$, then $\lambda_i = \gamma_i + 2\xi$, for $i = 1, 2, 3$. Since we must have $\lambda_2 = 0$, we have $\xi = -\frac{1}{2}\gamma_2$. □

Therefore, knowing H_L, we can fully recover the continuous homography matrix as $H = H_L - \frac{1}{2}\gamma_2 I$.

5.4.7 Decomposing the continuous homography matrix

We now address the task of decomposing the recovered $H = \widehat{\omega} + \frac{1}{d}vN^T$ into its motion and structure parameters $\{\omega, \frac{v}{d}, N\}$. The following constructive proof provides an algebraic technique for the recovery of motion and structure parameters.

Theorem 5.35 (Decomposition of continuous homography matrix). *Given a matrix $H \in \mathbb{R}^{3\times 3}$ in the form $H = \widehat{\omega} + \frac{1}{d}vN^T$, one can recover the motion and structure parameters $\{\widehat{\omega}, \frac{1}{d}v, N\}$ up to at most two physically possible solutions. There is a unique solution if $v = 0$, $v \times N = 0$, or $e_3^T v = 0$, where $e_3 = [0, 0, 1]^T$ is the optical axis.*

Proof. Compute the eigenvalue/eigenvector pairs of $H + H^T$ and denote them by $\{\lambda_i, u_i\}$, $i = 1, 2, 3$. If $\lambda_i = 0$ for $i = 1, 2, 3$, then we have $v = 0$ and $\widehat{\omega} = H$. In this case we cannot recover the normal of the plane N. Otherwise, if $\lambda_1 > 0$, and $\lambda_3 < 0$, then we have $v \times N \neq 0$. Let $\alpha = \|v/d\| > 0$, let $\tilde{v} = v/\sqrt{\alpha}$ and $\tilde{N} = \sqrt{\alpha}N$, and let $\beta = \tilde{v}^T\tilde{N}$. According to Lemma 5.33, the eigenvalue/eigenvector pairs of $H + H^T$ are given by

$$\begin{aligned} \lambda_1 &= \beta + \alpha > 0, \qquad u_1 = \tfrac{1}{\|\tilde{v}+\tilde{N}\|}(\tilde{v} + \tilde{N}), \\ \lambda_3 &= \beta - \alpha < 0, \qquad u_3 = \pm\tfrac{1}{\|\tilde{v}-\tilde{N}\|}(\tilde{v} - \tilde{N}). \end{aligned} \tag{5.81}$$

Then $\alpha = \frac{1}{2}(\lambda_1 - \lambda_3)$. It is easy to check that $\|\tilde{v} + \tilde{N}\|^2 = 2\lambda_1$, $\|\tilde{v} - \tilde{N}\|^2 = -2\lambda_3$. Together with (5.81), we have two solutions (due to the two possible signs for u_3):

$$\begin{aligned} \tilde{v}_1 &= \tfrac{1}{2}\big(\sqrt{2\lambda_1}\, u_1 + \sqrt{-2\lambda_3}\, u_3\big), & \tilde{v}_2 &= \tfrac{1}{2}\big(\sqrt{2\lambda_1}\, u_1 - \sqrt{-2\lambda_3}\, u_3\big), \\ \tilde{N}_1 &= \tfrac{1}{2}\big(\sqrt{2\lambda_1}\, u_1 - \sqrt{-2\lambda_3}\, u_3\big), & \tilde{N}_2 &= \tfrac{1}{2}\big(\sqrt{2\lambda_1}\, u_1 + \sqrt{-2\lambda_3}\, u_3\big), \\ \widehat{\omega}_1 &= H - \tilde{v}_1\tilde{N}_1^T, & \widehat{\omega}_2 &= H - \tilde{v}_2\tilde{N}_2^T. \end{aligned}$$

In the presence of noise, the estimate of $\widehat{\omega} = H - \tilde{v}\tilde{N}^T$ is not necessarily an element in $so(3)$. In algorithms, one may take its skew-symmetric part,

$$\widehat{\omega} = \frac{1}{2}\Big((H - \tilde{v}\tilde{N}^T) - (H - \tilde{v}\tilde{N}^T)^T\Big).$$

There is another sign ambiguity, since $(-\tilde{v})(-\tilde{N})^T = \tilde{v}\tilde{N}^T$. This sign ambiguity leads to a total of four possible solutions for decomposing H back to $\{\widehat{\omega}, \frac{1}{d}v, N\}$ given in Table 5.2.

Solution 1	$\frac{1}{d}v_1 = \sqrt{\alpha}\tilde{v}_1$ $N_1 = \frac{1}{\sqrt{\alpha}}\tilde{N}_1$ $\widehat{\omega}_1 = H - \tilde{v}_1\tilde{N}_1^T$	Solution 3	$\frac{1}{d}v_3 = -\frac{1}{d}v_1$ $N_3 = -N_1$ $\widehat{\omega}_3 = \widehat{\omega}_1$
Solution 2	$\frac{1}{d}v_2 = \sqrt{\alpha}\tilde{v}_2$ $N_2 = \frac{1}{\sqrt{\alpha}}\tilde{N}_2$ $\widehat{\omega}_2 = H - \tilde{v}_2\tilde{N}_2^T$	Solution 4	$\frac{1}{d}v_4 = -\frac{1}{d}v_2$ $v_4 = -N_2$ $\widehat{\omega}_4 = \widehat{\omega}_2$

Table 5.2. Four solutions for continuous planar homography decomposition. Here α is computed as before as $\alpha = \frac{1}{2}(\lambda_1 - \lambda_3)$.

In order to reduce the number of physically possible solutions, we impose the positive depth constraint: since the camera can only see points that are in front of it, we must have $N^T e_3 > 0$. Therefore, if solution 1 is the correct one, this constraint will eliminate solution 3 as being physically impossible. If $v^T e_3 \neq 0$, one of solutions 2 or 4 will be eliminated, whereas if $v^T e_3 = 0$, both solutions 2 and 4 will be eliminated. For the case that $v \times N = 0$, it is easy to see that solutions 1 and 2 are equivalent, and that imposing the positive depth constraint leads to a unique solution. □

Despite the fact that as in the discrete case, there is a close relationship between the continuous epipolar constraint and continuous homography, we will not develop the details here. Basic intuition and necessary technical tools have already been established in this chapter, and at this point interested readers may finish that part of the story with ease, or more broadly, apply these techniques to solve other special problems that one may encounter in real-world applications.

We summarize Sections 5.4.6 and 5.4.7 by presenting the continuous four-point Algorithm 5.4 for motion estimation from a planar scene.

5.5 Summary

Given corresponding points in two images $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ of a point p, or, in continuous time, optical flow $(\boldsymbol{u}, \boldsymbol{x})$, we summarize the constraints and relations between the image data and the unknown motion parameters in Table 5.3.

Despite the similarity between the discrete and the continuous case, one must be aware that there are indeed important subtle differences between these two cases, since the differentiation with respect to time t changes the algebraic relation between image data and unknown motion parameters.

In the presence of noise, the motion recovery problem in general becomes a problem of minimizing a cost function associated with statistical optimality or geometric error criteria subject to the above constraints. Once the camera motion is recovered, an overall 3-D reconstruction of both the camera motion and scene structure can be obtained up to a global scaling factor.

Algorithm 5.4 (The continuous four-point algorithm for a planar scene).

For a given set of optical flow vectors $(\boldsymbol{u}^j, \boldsymbol{x}^j)$, $j = 1, 2, \ldots, n$ $(n \geq 4)$, of points on a plane $N^T \boldsymbol{X} = d$, this algorithm finds $\{\widehat{\omega}, \frac{1}{d}v, N\}$ that solves

$$\widehat{\boldsymbol{x}^j}^T \left(\widehat{\omega} + \frac{1}{d} v N^T \right) \boldsymbol{x}^j = \widehat{\boldsymbol{x}^j} \boldsymbol{u}^j, \quad j = 1, 2, \ldots, n.$$

1. **Compute a first approximation of the continuous homography matrix**
   Construct the matrix $\chi = [\boldsymbol{a}^1, \boldsymbol{a}^2, \ldots, \boldsymbol{a}^n]^T \in \mathbb{R}^{3n \times 9}, B = [\boldsymbol{b}^{1T}, \boldsymbol{b}^{2T}, \ldots, \boldsymbol{b}^{nT}]^T \in \mathbb{R}^{3n}$ from the optical flow $(\boldsymbol{u}^j, \boldsymbol{x}^j)$, where $\boldsymbol{a}^j = \boldsymbol{x}^j \otimes \widehat{\boldsymbol{x}^j} \in \mathbb{R}^{9 \times 3}$ and $\boldsymbol{b} = \widehat{\boldsymbol{x}} \boldsymbol{u} \in \mathbb{R}^3$. Find the vector $H_L^s \in \mathbb{R}^9$ as
   $$H_L^s = \chi^\dagger B,$$
   where $\chi^\dagger \in \mathbb{R}^{9 \times 3n}$ is the pseudo-inverse of χ. Unstack H_L^s to obtain the 3×3 matrix H_L.
2. **Normalization of the continuous homography matrix**
   Compute the eigenvalue values $\{\gamma_1, \gamma_2, \gamma_3\}$ of the matrix $H_L^T + H_L$ and normalize it as
   $$H = H_L - \frac{1}{2}\gamma_2 I.$$
3. **Decomposition of the continuous homography matrix**
   Compute the eigenvalue decomposition of
   $$H^T + H = U \Lambda U^T$$
   and compute the four solutions for a decomposition $\{\widehat{\omega}, \frac{1}{d}v, N\}$ as in the proof of Theorem 5.35. Select the two physically possible ones by imposing the positive depth constraint $N^T e_3 > 0$.

5.6 Exercises

Exercise 5.1 (Linear equation). Solve $x \in \mathbb{R}^n$ from the linear equation

$$Ax = b,$$

where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. In terms of conditions on the matrix A and vector b, describe when a solution exists and when it is unique. In case the solution is not unique, describe the entire solution set.

Exercise 5.2 (Properties of skew-symmetric matrices).

1. Prove Lemma 5.4.
2. Prove Lemma 5.24.

Exercise 5.3 (Skew-symmetric matrix continued). Given a vector $T \in \mathbb{R}^3$ with unit length, i.e. $\|T\| = 1$, show that:

1. The identity holds: $\widehat{T}^T \widehat{T} = \widehat{T} \widehat{T}^T = I - TT^T$ (note that the superscript ^{T} stands for matrix transpose).

	Epipolar constraint	(Planar) homography
Discrete motion	$\boldsymbol{x}_2^T\widehat{T}R\boldsymbol{x}_1 = 0$	$\widehat{\boldsymbol{x}_2}(R + \frac{1}{d}TN^T)\boldsymbol{x}_1 = 0$
Matrices	$E = \widehat{T}R$	$H = R + \frac{1}{d}TN^T$
Relation	$\exists v \in \mathbb{R}^3,\ H = \widehat{T}^T E + Tv^T$	
Continuous motion	$\boldsymbol{x}^T\widehat{\omega}\widehat{v}\boldsymbol{x} + \boldsymbol{u}^T\widehat{v}\boldsymbol{x} = 0$	$\widehat{\boldsymbol{x}}(\widehat{\omega} + \frac{1}{d}vN^T)\boldsymbol{x} = \widehat{\boldsymbol{u}}\boldsymbol{x}$
Matrices	$E = \begin{bmatrix} \frac{1}{2}(\widehat{\omega}\widehat{v} + \widehat{v}\widehat{\omega}) \\ \widehat{v} \end{bmatrix}$	$H = \widehat{\omega} + \frac{1}{d}vN^T$
Linear algorithms	8 points	4 points
Decomposition	1 solution	2 solutions

Table 5.3. Here the number of points is required by corresponding linear algorithms, and we count only the number of physically possible solutions from corresponding decomposition algorithms *after* applying the positive depth constraint.

2. Explain the effect of multiplying a vector $u \in \mathbb{R}^3$ by the matrix $P = I - TT^T$. Show that $P^n = P$ for any integer n.
3. Show that $\widehat{T}^T\widehat{T}\widehat{T} = \widehat{T}\widehat{T}^T\widehat{T} = \widehat{T}$. Explain geometrically why this is true.
4. How do the above statements need to be changed if the vector T is not of unit length?

Exercise 5.4 (A rank condition for the epipolar constraint). Show that $\boldsymbol{x}_2^T\widehat{T}R\boldsymbol{x}_1 = 0$ if and only if

$$\text{rank}\,[\widehat{\boldsymbol{x}_2}R\boldsymbol{x}_1,\ \ \widehat{\boldsymbol{x}_2}T] \leq 1.$$

Exercise 5.5 (Parallel epipolar lines). Explain under what conditions the family of epipolar lines in at least one of the image planes will be parallel to each other. Where is the corresponding epipole (in terms of its homogeneous coordinates)?

Exercise 5.6 (Essential matrix for planar motion). Suppose we know that the camera always moves on a plane, say the XY plane. Show that:

1. The essential matrix $E = \widehat{T}R$ is of the special form

$$E = \begin{bmatrix} 0 & 0 & a \\ 0 & 0 & b \\ c & d & 0 \end{bmatrix}, \quad a, b, c, d \in \mathbb{R}. \tag{5.82}$$

2. Without using the SVD-based decomposition introduced in this chapter, find a solution to (R, T) in terms of a, b, c, d.

Exercise 5.7 (Rectified essential matrix). Suppose that using the linear algorithm, you obtain an essential matrix E of the form

$$E = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & a \\ 0 & -a & 0 \end{bmatrix}, \quad a \in \mathbb{R}. \tag{5.83}$$

What type of motion (R, T) does the camera undergo? How many solutions exist exactly?

Exercise 5.8 (Triangulation). Given two images $\boldsymbol{x}_1, \boldsymbol{x}_2$ of a point p together with the relative camera motion (R, T), $\boldsymbol{X}_2 = R\boldsymbol{X}_1 + T$:

1. express the depth of p with respect to the first image, i.e. λ_1 in terms of $\boldsymbol{x}_1, \boldsymbol{x}_2$, and (R, T);
2. express the depth of p with respect to the second image, i.e. λ_2 in terms of $\boldsymbol{x}_1, \boldsymbol{x}_2$, and (R, T).

Exercise 5.9 (Rotational motion). Assume that the camera undergoes pure rotational motion; i.e. it rotates around its center. Let $R \in SO(3)$ be the rotation of the camera and $\omega \in so(3)$ be the angular velocity. Show that in this case, we have:

1. discrete case: $\boldsymbol{x}_2^T \widehat{T} R \boldsymbol{x}_1 \equiv 0, \quad \forall T \in \mathbb{R}^3$;
2. continuous case: $\boldsymbol{x}^T \widehat{\omega}\widehat{v}\boldsymbol{x} + \boldsymbol{u}^T \widehat{v}\boldsymbol{x} \equiv 0, \quad \forall v \in \mathbb{R}^3$.

Exercise 5.10 (Projection onto $O(3)$). Given an arbitrary 3×3 matrix $M \in \mathbb{R}^{3\times 3}$ with positive singular values, find the orthogonal matrix $R \in O(3)$ such that the error $\|R - M\|_f^2$ is minimized. Is the solution unique? Note: Here we allow $\det(R) = \pm 1$.

Exercise 5.11 (Four motions related to an epipolar constraint). Suppose $E = \widehat{T}R$ is a solution to the epipolar constraint $\boldsymbol{x}_2^T E \boldsymbol{x}_1 = 0$. Then $-E$ is also an essential matrix, which obviously satisfies the same epipolar constraint (for given corresponding images).

1. Explain geometrically how these four motions are related. [Hint: Consider a pure translation case. If R is a rotation about T by an angle π, then $\widehat{T}R = -\widehat{T}$, which is in fact the *twisted pair* ambiguity.]
2. Show that in general, for three out of the four solutions, the equation $\lambda_2 \boldsymbol{x}_2 = \lambda_1 R \boldsymbol{x}_1 + T$ will yield either negative λ_1 or negative λ_2 or both. Hence only one solution satisfies the positive depth constraint.

Exercise 5.12 (Geometric distance to an epipolar line). Given two image points $\boldsymbol{x}_1, \tilde{\boldsymbol{x}}_2$ with respect to camera frames with their relative motion (R, T), show that the geometric distance d_2 defined in Figure 5.7 is given by the formula

$$d_2^2 = \frac{(\tilde{\boldsymbol{x}}_2^T \widehat{T} R \boldsymbol{x}_1)^2}{\|\widehat{e}_3 \widehat{T} R \boldsymbol{x}_1\|^2},$$

where $e_3 = [0, 0, 1]^T \in \mathbb{R}^3$.

Exercise 5.13 (A six-point algorithm). In this exercise, we show how to use some of the (algebraic) structure of the essential matrix to reduce the number of matched pairs of points from 8 to 6.

1. Show that if a matrix E is an essential matrix, then it satisfies the identity
$$EE^T E = \frac{1}{2}\text{trace}(EE^T)E.$$
2. Show that the dimension of the space of matrices $\{F\} \subset \mathbb{R}^{3\times 3}$ that satisfy the epipolar constraints
$$(\boldsymbol{x}_2^j)^T F \boldsymbol{x}_1^j = 0, \quad j = 1, 2, \ldots, 6,$$

is three. Hence the essential matrix E can be expressed as a linear combination $E = \alpha_1 F_1 + \alpha_2 F_2 + \alpha_3 F_3$ for some linearly independent matrices F_1, F_2, F_3 that satisfy the above equations.

3. To further determine the coefficients $\alpha_1, \alpha_2, \alpha_3$, show that the identity in (a) gives nine scalar equations linearly in the nine unknowns $\{\alpha_1^i \alpha_2^j \alpha_3^k\}$, $i + j + k = 3$, $0 \leq i, j, k \leq 3$. (Why nine?) Hence, the essential matrix E can be determined from six pairs of matched points.

Exercise 5.14 (Critical surfaces). To have a unique solution (up to a scalar factor), it is very important for the points considered in the above six-point or eight-point algorithms to be in general position. If a (dense) set of points whose images allow at least two distinct essential matrices, we say that they are "critical," Let $\boldsymbol{X} \in \mathbb{R}^3$ be coordinates of such a point and (R, T) be the motion of a camera. Let $\boldsymbol{x}_1 \sim \boldsymbol{X}$ and $\boldsymbol{x}_2 \sim (R\boldsymbol{X} + T)$ be two images of the point.

1. Show that if
$$(R\boldsymbol{X} + T)^T \widehat{T'} R' \boldsymbol{X} = 0,$$
then
$$\boldsymbol{x}_2^T \widehat{T} R \boldsymbol{x}_1 = 0, \quad \boldsymbol{x}_2^T \widehat{T'} R' \boldsymbol{x}_1 = 0.$$

2. Show that for points $\boldsymbol{X} \in \mathbb{R}^3$ that satisfy the equation $(R\boldsymbol{X} + T)^T \widehat{T'} R' \boldsymbol{X} = 0$, their homogeneous coordinates $\bar{\boldsymbol{X}} = [X, 1]^T \in \mathbb{R}^4$ satisfy the quadratic equation
$$\bar{\boldsymbol{X}}^T \begin{bmatrix} R^T \widehat{T'} R' + R'^T \widehat{T'}^T R & R'^T \widehat{T'}^T T \\ T^T \widehat{T'} R' & 0 \end{bmatrix} \bar{\boldsymbol{X}} = 0.$$
This quadratic surface is denoted by $C_1 \subset \mathbb{R}^3$ and is called a *critical surface*. So no matter how many points one chooses on such a surface, their two corresponding images always satisfy epipolar constraints for at least two different essential matrices.

3. Symmetrically, points defined by the equation $(R'\boldsymbol{X} + T')^T \widehat{T} R \boldsymbol{X} = 0$ will have similar properties. This gives another quadratic surface,
$$C_2 : \bar{\boldsymbol{X}}^T \begin{bmatrix} R'^T \widehat{T} R + R^T \widehat{T}^T R' & R^T \widehat{T}^T T' \\ T'^T \widehat{T} R & 0 \end{bmatrix} \bar{\boldsymbol{X}} = 0.$$
Argue that a set of points on the surface C_1 observed from two vantage points related by (R, T) could be interpreted as a corresponding set of points on the surface C_2 observed from two vantage points related by (R', T').

Exercise 5.15 (Estimation of the homography). We say that two images are related by a *homography* if the homogeneous coordinates of the two images $\boldsymbol{x}_1, \boldsymbol{x}_2$ of every point satisfy
$$\boldsymbol{x}_2 \sim H \boldsymbol{x}_1$$
for some nonsingular matrix $H \in \mathbb{R}^{3 \times 3}$. Show that in general one needs four pairs of $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ to determine the matrix H (up to a scalar factor).

Exercise 5.16 Under a homography $H \in \mathbb{R}^{3\times 3}$ from $\mathbb{R}^2$ to $\mathbb{R}^2$, a standard unit square with the homogeneous coordinates for the four corners

$$(0,0,1),\ (1,0,1),\ (1,1,1),\ (0,1,1)$$

is mapped to

$$(6,5,1),\ (4,3,1),\ (6,4.5,1),\ (10,8,1),$$

respectively. Determine the matrix H with its last entry H_{33} normalized to 1.

Exercise 5.17 (Epipolar line homography from an essential matrix). From the geometric interpretation of epipolar lines in Figure 5.2, we know that there is a one-to-one map between the family of epipolar lines $\{\boldsymbol{\ell}_1\}$ in the first image plane (through the epipole $\boldsymbol{e}_1$) and the family of epipolar lines $\{\boldsymbol{\ell}_2\}$ in the second. Suppose that the essential matrix E is known. Show that this map is in fact a homography. That is, there exists a nonsingular matrix $H \in \mathbb{R}^{3\times 3}$ such that

$$\boldsymbol{\ell}_2 \sim H\boldsymbol{\ell}_1$$

for any pair of corresponding epipolar lines $(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2)$. Find an explicit form for H in terms of E.

Exercise 5.18 (Homography with respect to the second camera frame). In the chapter, we have learned that for a transformation $\boldsymbol{X}_2 = R\boldsymbol{X}_1 + T$ on a plane $N^T\boldsymbol{X}_1 = 1$ (expressed in the first camera frame), we have a homography $H = R + TN^T$ such that $\boldsymbol{x}_2 \sim H\boldsymbol{x}_1$ relates the two images of the plane.

1. Now switch roles of the first and the second camera frames and show that the new homography matrix becomes

$$\tilde{H} = \left(R^T + \frac{-R^T T}{1 + N^T R^T T} N^T R^T\right). \tag{5.84}$$

2. What is the relationship between H and $\tilde{H}$? Provide a formal proof to your answer. Explain why this should be expected.

Exercise 5.19 (Two physically possible solutions for the homography decomposition). Let us study in the nature of the two physically possible solutions for the homography decomposition. Without loss of generality, suppose that the true homography matrix is $H = I + ab^T$ with $\|a\| = 1$.

1. Show that $R' = -I + 2aa^T$ is a rotation matrix.
2. Show that $H' = R' + (-a)(b + 2a)^T$ is equal to $-H$.
3. Since $(H')^T H' = H^T H$, conclude that both $\{I, a, b\}$ and $\{R', -a, (b + 2a)\}$ are solutions from the homography decomposition of H.
4. Argue that, under certain conditions on the relationship between a and b, the second solution is also physically possible.
5. What is the geometric relationship between these two solutions? Draw a figure to illustrate your answer.

Exercise 5.20 (Various expressions for the image motion field). In the continuous-motion case, suppose that the camera motion is (ω, v), and $\boldsymbol{u} = \dot{\boldsymbol{x}}$ is the velocity of the image $\boldsymbol{x}$ of a point $\boldsymbol{X} = [X, Y, Z]^T$ in space. Show that:

1. For a spherical perspective projection; i.e. $\lambda = \|X\|$, we have
$$\boldsymbol{u} = -\widehat{\boldsymbol{x}}\omega + \frac{1}{\lambda}\widehat{\boldsymbol{x}}^2 v. \tag{5.85}$$
2. For a planar perspective projection; i.e. $\lambda = Z$, we have
$$\boldsymbol{u} = (-\widehat{\boldsymbol{x}} + \boldsymbol{x}e_3^T\widehat{\boldsymbol{x}})\omega + \frac{1}{\lambda}(I - \boldsymbol{x}e_3^T)v, \tag{5.86}$$
or in coordinates,
$$\begin{bmatrix}\dot{x}\\ \dot{y}\end{bmatrix} = \begin{bmatrix}-xy & x^2 & -y\\ -(1+y^2) & xy & x\end{bmatrix}\omega + \frac{1}{\lambda}\begin{bmatrix}1 & 0 & -x\\ 0 & 1 & -y\end{bmatrix}v. \tag{5.87}$$
3. Show that in the planar perspective case, equation (5.76) is equivalent to
$$\boldsymbol{u} = (I - \boldsymbol{x}e_3^T)H\boldsymbol{x}. \tag{5.88}$$
From this equation, discuss under what conditions the motion field for a planar scene is an affine function of the image coordinates; i.e.
$$\boldsymbol{u} = A\boldsymbol{x}, \tag{5.89}$$
where A is a constant 3×3 affine matrix that does not depend on the image point $\boldsymbol{x}$.

Exercise 5.21 (Programming: implementation of (discrete) eight-point algorithm). Implement a version of the three-step pose estimation algorithm for two views. Your Matlab code should be responsible for

- Initialization: Generate a set of n (≥ 8) 3-D points; generate a rigid-body motion (R, T) between two camera frames and project (the coordinates of) the points (relative to the camera frame) onto the image plane correctly. Here you may assume that the focal length is 1. This step will give you corresponding images as input to the algorithm.
- Motion Recovery: using the corresponding images and the algorithm to compute the motion $(\tilde{R}, \tilde{T})$ and compare it to the ground truth (R, T).

After you get the correct answer from the above steps, here are a few suggestions for you to try with the algorithm (or improve it):

- A more realistic way to generate these 3-D points is to make sure that they are all indeed "in front of" the image plane before and after the camera moves.
- Systematically add some noise to the projected images and see how the algorithm responds. Try different camera motions and different layouts of the points in 3-D.
- Finally, to make the algorithm fail, take all the 3-D points from some plane in front of the camera. Run the program and see what you get (especially with some noise on the images).

Exercise 5.22 (Programming: implementation of the continuous eight-point algorithm). Implement a version of the four-step velocity estimation algorithm for optical flow.

- Initialization: Choose a set of n (≥ 8) 3-D points and a rigid-body velocity (ω, v). Correctly obtain the image $\boldsymbol{x}$ and compute the image velocity $\boldsymbol{u} = \dot{\boldsymbol{x}}$. You need to figure out how to compute $\boldsymbol{u}$ from (ω, v) and $\boldsymbol{X}$. Here you may assume that the focal length is 1. This step will give you images and their velocities as input to the algorithm.
- Motion Recovery: Use the algorithm to compute the motion $(\tilde{\omega}, \tilde{v})$ and compare it to the ground truth (ω, v).

5.A Optimization subject to the epipolar constraint

In this appendix, we will study the problem of minimizing the reprojection error (5.23) subject to the fact that the underlying unknowns must satisfy the epipolar constraint. This yields an optimal estimate, in the sense of least-squares, of camera motion between the two views.

Constraint elimination by Lagrange multipliers

Our goal here is, given $\tilde{\boldsymbol{x}}_i^j, i = 1, 2,\ j = 1, 2, \ldots, n$, to find

$$(\boldsymbol{x}^*, R^*, T^*) = \arg\min \phi(\boldsymbol{x}, R, T) \doteq \sum_{j=1}^{n} \sum_{i=1}^{2} \|\tilde{\boldsymbol{x}}_i^j - \boldsymbol{x}_i^j\|_2^2$$

subject to

$$\boldsymbol{x}_2^{jT} \widehat{T} R \boldsymbol{x}_1^j = 0, \quad \boldsymbol{x}_1^{jT} e_3 = 1, \quad \boldsymbol{x}_2^{jT} e_3 = 1, \quad j = 1, 2, \ldots, n. \tag{5.90}$$

Using Lagrange multipliers (Appendix C) $\lambda^j, \gamma^j, \eta^j$, we can convert the above minimization problem to an unconstrained minimization problem over $R \in SO(3), T \in \mathbb{S}^2, \boldsymbol{x}_1^j, \boldsymbol{x}_2^j, \lambda^j, \gamma^j, \eta^j$. Consider the Lagrangian function associated with this constrained optimization problem .

$$\min \sum_{j=1}^{n} \|\tilde{\boldsymbol{x}}_1^j - \boldsymbol{x}_1^j\|^2 + \|\tilde{\boldsymbol{x}}_2^j - \boldsymbol{x}_2^j\|^2 + \lambda^j \boldsymbol{x}_2^{jT} \widehat{T} R \boldsymbol{x}_1^j + \gamma^j (\boldsymbol{x}_1^{jT} e_3 - 1) + \eta^j (\boldsymbol{x}_2^{jT} e_3 - 1). \tag{5.91}$$

A necessary condition for the existence of a minimum is $\nabla L = 0$, where the derivative is taken with respect to $\boldsymbol{x}_1^j, \boldsymbol{x}_2^j, \lambda^j, \gamma^j, \eta^j$. Setting the derivative with respect to the Lagrange multipliers $\lambda^j, \gamma^j, \eta^j$ to zero returns the equality constraints, and setting the derivative with respect to $\boldsymbol{x}_1^j, \boldsymbol{x}_2^j$ to zero yields

$$\begin{aligned} 2(\tilde{\boldsymbol{x}}_1^j - \boldsymbol{x}_1^j) + \lambda^j R^T \widehat{T}^T \boldsymbol{x}_2^j + \gamma^j e_3 = 0, \\ 2(\tilde{\boldsymbol{x}}_2^j - \boldsymbol{x}_2^j) + \lambda^j \widehat{T} R \boldsymbol{x}_1^j + \eta^j e_3 = 0. \end{aligned}$$

Simplifying these equations by premultiplying both by the matrix $\widehat{e}_3^T \widehat{e}_3$, we obtain

$$\begin{aligned} \boldsymbol{x}_1^j &= \tilde{\boldsymbol{x}}_1^j - \tfrac{1}{2} \lambda^j \widehat{e}_3^T \widehat{e}_3 R^T \widehat{T}^T \boldsymbol{x}_2^j, \\ \boldsymbol{x}_2^j &= \tilde{\boldsymbol{x}}_2^j - \tfrac{1}{2} \lambda^j \widehat{e}_3^T \widehat{e}_3 \widehat{T} R \boldsymbol{x}_1^j. \end{aligned} \tag{5.92}$$

Together with $\boldsymbol{x}_2^{jT}\widehat{T}R\boldsymbol{x}_1^j = 0$, we may solve for the Lagrange multipliers λ^j in different expressions,[22]

$$\lambda^j = \frac{2(\boldsymbol{x}_2^{jT}\widehat{T}R\tilde{\boldsymbol{x}}_1^j + \tilde{\boldsymbol{x}}_2^{jT}\widehat{T}R\boldsymbol{x}_1^j)}{\boldsymbol{x}_1^{jT}R^T\widehat{T}^T\widehat{e}_3^T\widehat{e}_3\widehat{T}R\boldsymbol{x}_1^j + \boldsymbol{x}_2^{jT}\widehat{T}R\widehat{e}_3^T\widehat{e}_3R^T\widehat{T}^T\boldsymbol{x}_2^j} \tag{5.93}$$

or

$$\lambda^j = \frac{2\boldsymbol{x}_2^{jT}\widehat{T}R\tilde{\boldsymbol{x}}_1^j}{\boldsymbol{x}_1^{jT}R^T\widehat{T}^T\widehat{e}_3^T\widehat{e}_3\widehat{T}R\boldsymbol{x}_1^j} = \frac{2\tilde{\boldsymbol{x}}_2^{jT}\widehat{T}R\boldsymbol{x}_1^j}{\boldsymbol{x}_2^{jT}\widehat{T}R\widehat{e}_3^T\widehat{e}_3R^T\widehat{T}^T\boldsymbol{x}_2^j}. \tag{5.94}$$

Substituting (5.92) and (5.93) into the least-squares cost function of equation (5.91), we obtain

$$\phi(\boldsymbol{x}, R, T) = \sum_{j=1}^{n} \frac{(\boldsymbol{x}_2^{jT}\widehat{T}R\tilde{\boldsymbol{x}}_1^j + \tilde{\boldsymbol{x}}_2^{jT}\widehat{T}R\boldsymbol{x}_1^j)^2}{\|\widehat{e}_3\widehat{T}R\boldsymbol{x}_1^j\|^2 + \|\boldsymbol{x}_2^{jT}\widehat{T}R\widehat{e}_3^T\|^2}. \tag{5.95}$$

If one uses instead (5.92) and (5.94), one gets

$$\phi(\boldsymbol{x}, R, T) = \sum_{j=1}^{n} \frac{(\tilde{\boldsymbol{x}}_2^{jT}\widehat{T}R\boldsymbol{x}_1^j)^2}{\|\widehat{e}_3\widehat{T}R\boldsymbol{x}_1^j\|^2} + \frac{(\boldsymbol{x}_2^{jT}\widehat{T}R\tilde{\boldsymbol{x}}_1^j)^2}{\|\boldsymbol{x}_2^{jT}\widehat{T}R\widehat{e}_3^T\|^2}. \tag{5.96}$$

These expressions for ϕ can finally be minimized with respect to (R, T) as well as $\boldsymbol{x} = \{\boldsymbol{x}_i^j\}$. In doing so, however, one has to make sure that the unknwns are constrained so that $R \in SO(3)$ and $T \in \mathbb{S}^2$ are explicitly enforced. In Appendix C we discuss methods for minimizing a function with unknowns in spaces like $SO(3) \times \mathbb{S}^2$, that can be used to minimize $\phi(\boldsymbol{x}, R, T)$ once $\boldsymbol{x}$ is known. Since $\boldsymbol{x}$ is *not* known, one can set up an *alternating minimization scheme where an initial approximation of $\boldsymbol{x}$ is used to estimate an approximation of (R, T), which is used, in turn, to update the estimates of $\boldsymbol{x}$.* It can be shown that each such iteration decreases the cost function, and therefore convergence to a local extremum is guaranteed, since the cost function is bounded below by zero. The overall process is described in Algorithm 5.5. As we mentioned before, this is equivalent to the so-called *bundle adjustment* for the two-view case, that is the direct minimization of the reprojection error with respect to all unknowns. Equivalence is intended in the sense that, at the optimum, the two solutions coincide.

Structure triangulation

In step 3 of Algorithm 5.5, for each pair of images $(\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2)$ and a fixed (R, T), $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ can be computed by minimizing the same reprojection error function $\phi(\boldsymbol{x}) = \|\tilde{\boldsymbol{x}}_1 - \boldsymbol{x}_1\|^2 + \|\tilde{\boldsymbol{x}}_2 - \boldsymbol{x}_2\|^2$ for each pair of image points. Assuming that the notation is the same as in Figure 5.9, let $\boldsymbol{\ell}_2 \in \mathbb{R}^3$ be the normal vector (of unit length) to the epipolar plane spanned by $(\boldsymbol{x}_2, \boldsymbol{e}_2)$.[23] Given such an $\boldsymbol{\ell}_2$, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$

[22] Since we have multiple equations to solve for one unknown λ^j, the redundancy gives rise to different expressions depending on which equation in (5.92) is used.

[23] $\boldsymbol{\ell}_2$ can also be interpreted as the coimage of the epipolar line in the second image, but here we do not use that interpretation.

Algorithm 5.5 (Optimal triangulation).

1. **Initialization**
   Initialize $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ as $\tilde{\boldsymbol{x}}_1$ and $\tilde{\boldsymbol{x}}_2$, respectively. Also initialize (R, T) with the pose initialized by the solution from the eight-point linear algorithm.
2. **Pose estimation**
   For $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ computed from the previous step, update (R, T) by minimizing the reprojection error $\phi(\boldsymbol{x}, R, T)$ given in its unconstrained form (5.95) or (5.96).
3. **Structure triangulation**
   For each image pair $(\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2)$ and (R, T) computed from the previous step, solve for $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ that minimize the reprojection error $\phi(\boldsymbol{x}) = \|\boldsymbol{x}_1 - \tilde{\boldsymbol{x}}_1\|^2 + \|\boldsymbol{x}_2 - \tilde{\boldsymbol{x}}_2\|^2$.
4. Return to step 2 until the decrement in the value of ϕ is below a threshold.

are determined by

$$\boldsymbol{x}_1(\boldsymbol{\ell}_1) = \frac{\widehat{e}_3\boldsymbol{\ell}_1\boldsymbol{\ell}_1^T\widehat{e}_3^T\tilde{\boldsymbol{x}}_1 + \widehat{\boldsymbol{\ell}_1}^T\widehat{\boldsymbol{\ell}}_1 e_3}{e_3^T\widehat{\boldsymbol{\ell}_1}^T\widehat{\boldsymbol{\ell}}_1 e_3}, \quad \boldsymbol{x}_2(\boldsymbol{\ell}_2) = \frac{\widehat{e}_3\boldsymbol{\ell}_2\boldsymbol{\ell}_2^T\widehat{e}_3^T\tilde{\boldsymbol{x}}_2 + \widehat{\boldsymbol{\ell}_2}^T\widehat{\boldsymbol{\ell}}_2 e_3}{e_3^T\widehat{\boldsymbol{\ell}_2}^T\widehat{\boldsymbol{\ell}}_2 e_3},$$

where $\boldsymbol{\ell}_1 = R^T\boldsymbol{\ell}_2 \in \mathbb{R}^3$. Then the distance can be explicitly expressed as

$$\|\tilde{\boldsymbol{x}}_2 - \boldsymbol{x}_2\|^2 + \|\tilde{\boldsymbol{x}}_1 - \boldsymbol{x}_1\|^2 \;=\; \|\tilde{\boldsymbol{x}}_2\|^2 + \frac{\boldsymbol{\ell}_2^T A\boldsymbol{\ell}_2}{\boldsymbol{\ell}_2^T B\boldsymbol{\ell}_2} + \|\tilde{\boldsymbol{x}}_1\|^2 + \frac{\boldsymbol{\ell}_1^T C\boldsymbol{\ell}_1}{\boldsymbol{\ell}_1^T D\boldsymbol{\ell}_1},$$

where $A, B, C, D \in \mathbb{R}^{3\times 3}$ are defined as functions of $(\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2)$:

$$\begin{aligned} A &= I - (\widehat{e}_3\tilde{\boldsymbol{x}}_2\tilde{\boldsymbol{x}}_2^T\widehat{e}_3^T + \widehat{\tilde{\boldsymbol{x}}_2}\widehat{e}_3 + \widehat{e}_3\widehat{\tilde{\boldsymbol{x}}_2}), \quad B = \widehat{e}_3^T\widehat{e}_3, \\ C &= I - (\widehat{e}_3\tilde{\boldsymbol{x}}_1\tilde{\boldsymbol{x}}_1^T\widehat{e}_3^T + \widehat{\tilde{\boldsymbol{x}}_1}\widehat{e}_3 + \widehat{e}_3\widehat{\tilde{\boldsymbol{x}}_1}), \quad D = \widehat{e}_3^T\widehat{e}_3. \end{aligned} \tag{5.97}$$

Then the problem of finding the optimal $\boldsymbol{x}_1^*$ and $\boldsymbol{x}_2^*$ becomes a problem of finding the normal vector $\boldsymbol{\ell}_2^*$ that minimizes the function of a sum of two *singular Rayleigh quotients*:

$$\min_{\boldsymbol{\ell}_2^T T=0, \boldsymbol{\ell}_2^T\boldsymbol{\ell}_2=1} V(\boldsymbol{\ell}_2) \;=\; \frac{\boldsymbol{\ell}_2^T A\boldsymbol{\ell}_2}{\boldsymbol{\ell}_2^T B\boldsymbol{\ell}_2} + \frac{\boldsymbol{\ell}_2^T RCR^T\boldsymbol{\ell}_2}{\boldsymbol{\ell}_2^T RDR^T\boldsymbol{\ell}_2}. \tag{5.98}$$

This is an optimization problem on the unit circle $\mathbb{S}^1$ in the plane orthogonal to the (epipole) vector $\boldsymbol{e}_2(\sim T)$.[24] If $N_1, N_2 \in \mathbb{R}^3$ are vectors such that $(\boldsymbol{e}_2, N_1, N_2)$ form an orthonormal basis of $\mathbb{R}^3$ in the second camera frame, then $\boldsymbol{\ell}_2 = \cos(\theta)N_1 + \sin(\theta)N_2$ with $\theta \in \mathbb{R}$. We need only to find θ^* that minimizes the function $V(\boldsymbol{\ell}_2(\theta))$. From the geometric interpretation of the optimal solution, we also know that the global minimum θ^* should lie between two values: θ_1 and θ_2 such that $\boldsymbol{\ell}_2(\theta_1)$ and $\boldsymbol{\ell}_2(\theta_2)$ correspond to normal vectors of the two planes

[24] Therefore, geometrically, motion and structure recovery from n pairs of image correspondences is really an optimization problem on the space $SO(3) \times \mathbb{S}^2 \times \mathbb{T}^n$, where $\mathbb{T}^n$ is an n-torus, i.e. an n-fold product of $\mathbb{S}^1$.

spanned by $(\tilde{x}_2, e_2)$ and $(R\tilde{x}_1, e_2)$, respectively.[25] The problem now becomes a simple bounded minimization problem for a scalar function (in θ) and can be efficiently solved using standard optimization routines (such as "`fmin`" in Matlab or Newton's algorithm, described in Appendix C).

Historical notes

The origins of epipolar geometry can be dated back as early as the mid nineteenth century and appeared in the work of Hesse on studying the two-view geometry using seven points (see [Maybank and Faugeras, 1992] and references therein). Kruppa proved in 1913 that five points in general position are all one needs to solve the two-view problem up to a finite number of solutions [Kruppa, 1913]. Kruppa's proof was later improved in the work of [Demazure, 1988] where the actual number of solutions was proven, with a simpler proof given later by [Heyden and Sparr, 1999]. A constructive proof can be found in [Philip, 1996], and in particular, a linear algorithm is provided if there are six matched points, from which Exercise 5.13 was constructed. A more efficient five-point algorithm that enables real-time implementation has been recently implemented by [Nistér, 2003].

The eight-point and four-point algorithms

To our knowledge, the epipolar constraint first appeared in [Thompson, 1959]. The (discrete) eight-point linear algorithm introduced in this chapter is due to the work of [Longuet-Higgins, 1981] and [Huang and Faugeras, 1989], which sparked a wide interest in the structure from motion problem in computer vision and led to the development of numerous linear and nonlinear algorithms for motion estimation from two views. Early work on these subjects can be found in the books or manuscripts of [Faugeras, 1993, Kanatani, 1993b, Maybank, 1993, Weng et al., 1993b]. An improvement of the eight-point algorithm based on normalizing image coordinates was later given by [Hartley, 1997]. [Soatto et al., 1996] studied further the dynamical aspect of epipolar geometry and designed a Kalman filter on the manifold of essential matrices for dynamical motion estimation. We will study Kalman-filter-based approaches in Chapter 12.

The homography (discrete or continuous) between two images of a planar scene has been extensively studied and used in the computer vision literature. Early results on this subject can be found in [Subbarao and Waxman, 1985, Waxman and Ullman, 1985, Kanatani, 1985, Longuet-Higgins, 1986]. The four-point algorithm based on decomposing the homography matrix was first given by [Faugeras and Lustman, 1988]. A thorough discussion on the homography and the relationships between the two physically possible solutions in Theorem 5.19 can be found in [Weng et al., 1993b] and references therein. This chapter is a very

[25] If $\tilde{x}_1, \tilde{x}_2$ already satisfy the epipolar constraint, these two planes coincide.

concise summary and supplement to these early results in computer vision. In Chapter 9 we will see how the epipolar constraint and homography can be unified into a single type of constraint.

Critical surfaces

Regarding the criticality or ambiguity of the two-view geometry mentioned before (such as the critical surfaces), the interested reader may find more details in [Adiv, 1985, Longuet-Higgins, 1988, Maybank, 1993, Soatto and Brockett, 1998] or the book of [Faugeras and Luong, 2001]. More discussions on the criticality and degeneracy in camera calibration and multiple-view reconstruction can be found in later chapters.

Objective functions for estimating epipolar geometry

Many objective functions have been used in the computer vision literature for estimating the two-view epipolar geometry, such as "epipolar improvement" [Weng et al., 1993a], "normalized epipolar constraint" [Weng et al., 1993a, Luong and Faugeras, 1996, Zhang, 1998c], "minimizing the reprojection error" [Weng et al., 1993a], and "triangulation" [Hartley and Sturm, 1997]. The method presented in this chapter follows that of [Ma et al., 2001b].

As discussed in Section 5.A, there is no closed-form solution to an optimal motion and structure recovery problem if the reprojection error is chosen to be the objective since the problem involves solving algebraic equations of order six [Hartley and Sturm, 1997, Ma et al., 2001b]. The solution is typically found through iterative numerical schemes such as the ones described in Appendix C. It has, however, been shown by [Oliensis, 2001] that if one chooses to minimize the angle (not distance) between the measured $\tilde{x}$ and recovered x, a closed-form solution is available. Hence, solvability of a reconstruction problem does depend on the choice of objective function. In the multiple-view setting, minimizing reprojection error corresponds to a nonlinear optimization procedure [Spetsakis and Aloimonos, 1988], often referred to as "bundle adjustment," which we will discuss in Chapter 11.

The continuous motion case

The search for the continuous counterpart of the eight-point algorithm has produced many different versions in the computer vision literature due to its subtle difference from the discrete case. To our knowledge, the first algorithm was proposed in 1984 by [Zhuang and Haralick, 1984] with a simplified version given in [Zhuang et al., 1988]; and a first-order algorithm was given by [Waxman et al., 1987]. Other algorithms solved for rotation and translation separately using either numerical optimization techniques [Bruss and Horn, 1983] or linear subspace methods [Heeger and Jepson, 1992, Jepson and Heeger, 1993]. [Kanatani, 1993a] proposed a linear algorithm reformulating Zhuang's approach in terms of essential parameters and twisted flow. See [Tian et al., 1996] for some experimental comparisons of these methods, while analytical results on the sensitivity of two-view geometry can be found in [Daniilidis and Nagel, 1990, Spetsakis, 1994, Daniilidis and Spetsakis, 1997] and estimation bias study in the work of [Heeger and Jepson, 1992, Kanatani, 1993b]. [Fermüller et al., 1997] has further shown that the distortion induced on the structure from errors in the motion estimates is governed by the so-called Cremona transformation. The parallel development of the continuous eight-point algorithm presented in this chapter follows that of [Ma et al., 2000a], where the interested reader may also find a more detailed account of related bibliography and history. Besides the linear methods, a study of the (nonlinear) optimal solutions to the continuous motion case was given in [Chiuso et al., 2000].