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 of the cameras as well as the position of the points in space from their projection onto the two images.

Longuet-Higgins noticed that the coordinates of the projection of a point and the camera optical centers form a triangle (Figure 6.5), 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 solved 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 non-linear 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 algorithm of

Longuet-Higgins, although merely conceptual[1], allows us to introduce the basic ideas that will be revisited later in the chapter as well as in Part III and IV 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

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

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* (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 with origin o in the optical center and z-axis aligned with the optical axis. The relationship between 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 $\boldsymbol{X}_1 \in \mathbb{R}^3$ and $\boldsymbol{X}_2 \in \mathbb{R}^3$ the 3-D coordinates of a point p relative to the two camera frames, respectively, 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 can be written in terms of the image coordinates $\boldsymbol{x}_i$'s and the depths λ_i's as

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

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

In order to eliminate the depths λ_i's in the preceding equation, pre-multiply 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. Pre-multiplying 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 shown

Theorem 5.1 (Epipolar constraint). *Two images $\boldsymbol{x}_1, \boldsymbol{x}_2$ of a point p seen from two vantage points satisfy the following constraint*

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

where (R, T) is the relative pose (position and orientation) between the two camera reference frames.

The matrix

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

in the epipolar constraint (5.2) is called the *Essential matrix*. It encodes the relative pose between the two cameras. The epipolar constraint (5.2) can therefore be 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*, the reason of which will become clear in later chapters.

In addition to the preceding algebraic derivation, this constraint follows immediately from its geometric interpretation, as illustrated in Figure 6.5. The vector

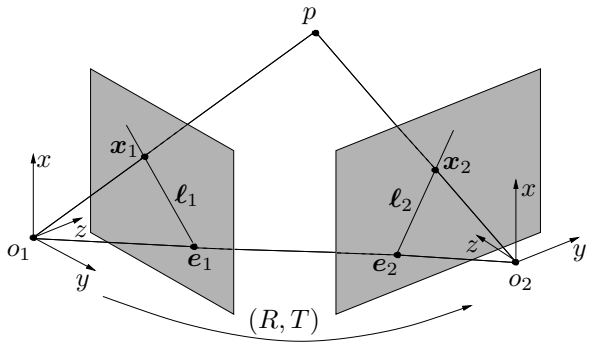

Figure 5.1. Two projections $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^3$ of a 3-D point p from two vantage points. The Euclidean transformation between the two vantage points is given by $(R, T) \in SE(3)$. The intersection of the line (o_1, o_2) with each image plane is the so-called *epipole*,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

connecting the first camera center o_1 and the point p, the vector connecting o_2 and p, and the vector connecting the two optical centers o_1 and o_2 clearly form a

triangle. Therefore, the three vectors lie in the same plane. Their triple product[2], 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) share the same direction. 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 to 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 $\boldsymbol{e}_1(\boldsymbol{e}_2)$ 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* epipolar line *of p. We usually use the normal vector $\boldsymbol{\ell}_1(\boldsymbol{\ell}_2)$ to the epipolar plane to denote this line.*[3]

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 $E = \widehat{T}R$ which defines an epipolar relation between two images $\boldsymbol{x}_1, \boldsymbol{x}_2$, we have:*

1. *The two epipoles $\boldsymbol{e}_1, \boldsymbol{e}_2 \in \mathbb{R}^3$, with respect to the 1^{st} and 2^{nd} camera frames respectively, are the left and right null space of E respectively*

$$\boldsymbol{e}_2^T E = 0, \quad E\boldsymbol{e}_1 = 0. \tag{5.3}$$

   That is, $\boldsymbol{e}_2 \sim T$ and $\boldsymbol{e}_1 \sim R^T T$. We recall that $\sim$ indicates equality up to scale.

2. *The (co-images 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*

$$\boldsymbol{\ell}_2 \sim E\boldsymbol{x}_1, \quad \boldsymbol{\ell}_1 \sim E^T\boldsymbol{x}_2 \quad \in \mathbb{R}^3 \tag{5.4}$$

[2] 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.

[3] Hence the vector $\boldsymbol{\ell}_1$ is in fact the co-image of the epipolar line.

where ℓ_1, ℓ_2 are in fact the normal vectors to the epipolar plane expressed with respect to the two camera frames, respectively.

3. *In each image, we have the relationship that both the projected point and the epipole lie on the epipolar line*

$$\ell_i^T e_i = 0, \quad \ell_i^T x_i = 0, \quad i = 1, 2. \tag{5.5}$$

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

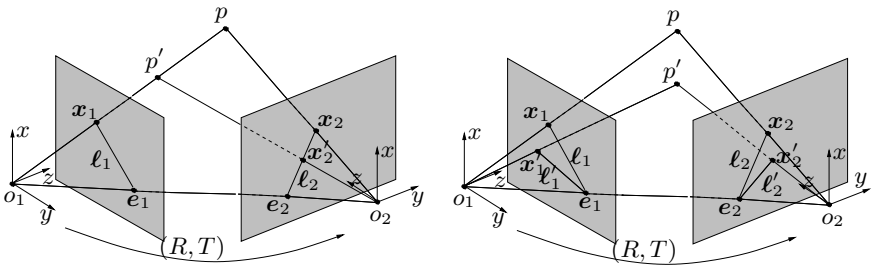

Figure 5.2. Left: the Essential matrix E associated to the epipolar constraint maps an image point x_1 in the first image to an epipolar line $\ell_2 = Ex_1$ in the second image – precise location of its corresponding image (x_2 or x_2') depends on where the 3-D point (p or p') lies on the ray (o_1, 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 (ℓ_1, ℓ_2) and (ℓ_1', ℓ_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 $E = \widehat{T}R \in \mathbb{R}^{3\times 3}$ 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 $\mathbb{R}^{3\times 3}$ called the *Essential space* and denote by $\mathcal{E}$

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

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 $K \in \mathbb{R}^{3\times 3}$, 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 $\mathbb{R}^{3\times 3}$, one may directly verify that these two linear maps agree on the basis $[1,0,0]^T, [0,1,0]^T$ or $[0,0,1]^T$ (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 non-zero 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*

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

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 $R_0 T = [0, 0, \|T\|]^T$. Denote this vector as $a \in \mathbb{R}^3$. 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 determinant can be ± 1.[4] We still need to prove that $U, V \in SO(3)$ (i.e. they have determinant $+1$) to establish the theorem. We already have $E = \widehat{T}R = R_0^T\widehat{a}R_0R$. Let $R_Z(\theta)$ be the matrix which represents a rotation around the Z-axis by an angle of θ radians, i.e. $R_Z(\theta) = e^{\widehat{e}_3\theta}$ 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 $\widehat{a} = R_Z(+\frac{\pi}{2})R_Z^T(+\frac{\pi}{2})\widehat{a} = R_Z(+\frac{\pi}{2})\ \text{diag}\{\|T\|, \|T\|, 0\}$. Therefore

$$E = \widehat{T}R = R_0^T R_Z\left(+\frac{\pi}{2}\right)\ \text{diag}\{\|T\|, \|T\|, 0\}R_0R.$$

So, in the SVD of $E = U\Sigma V^T$, we may choose $U = R_0^T R_Z(+\frac{\pi}{2})$ and $V^T = R_0R$. 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 $\Sigma = \text{diag}\{\sigma, \sigma, 0\}$, define $(R_1, T_1) \in SE(3)$ and $(R_2, T_2) \in SE(3)$ to be

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

[4] 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 rotation 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 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 non-zero 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 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 exists $\omega \in \mathbb{R}^3, \|\omega\| = 1$ and $\theta \in \mathbb{R}$ such that $R = e^{\widehat{\omega}\theta}$. 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 to 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. $\omega = \pm u$. R then 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) introduced in Chapter 2, we have

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

(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 follows. □

The following theorem shows exactly how many pairs of rotation and translation (R, T) can one 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 non-zero 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 $\widehat{T}_1 R_1 = \widehat{T}_2 R_2$. It 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 formulas give the two distinct solutions

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

□

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 how 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. ■

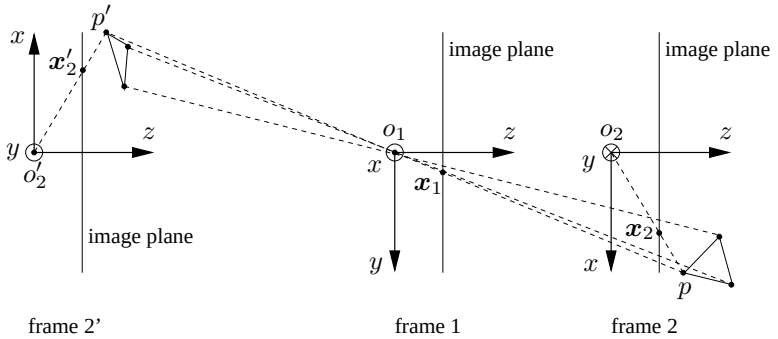

Figure 5.3. Two pairs of camera frames, i.e. $(1, 2)$ and $(1, 2')$, generate the same Essential matrix. The frame 2 and frame $2'$ differ by a translation and a 180^o rotation (a twist) around the Z-axis and the two pose pairs give rise to the same image coordinates. For the same set of image pairs $\boldsymbol{x}_1$ and $\boldsymbol{x}_2 = \boldsymbol{x}_2'$, 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(5.9).

Although the linear algorithm that we propose here is suboptimal when the measurements are corrupted by noise, it is important for it illustrates that the geometric structure of the space of Essential matrices is at the heart of the problem of reconstruction from two views. 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). When the entries of this 3×3 matrix are denoted as

$$E = \begin{bmatrix} e_1 & e_2 & e_3 \\ e_4 & e_5 & e_6 \\ e_7 & e_8 & e_9 \end{bmatrix} \tag{5.10}$$

and arrayed in 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)

$$E^s \doteq [e_1, e_4, e_7, e_2, e_5, e_8, e_3, e_6, e_9]^T \quad \in \mathbb{R}^9.$$

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 $\boldsymbol{x}_2$ as

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

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 j^{th} 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 scale, 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 noisy (as the images shown), 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. Another condition to be aware of is when the rank of χ is less than 8 regardless the number of points used, allowing for multiple solutions to equation (5.14). This can happen when the feature points are not in "general position", for example when they all lie in a plane (as we will soon see in the next section).

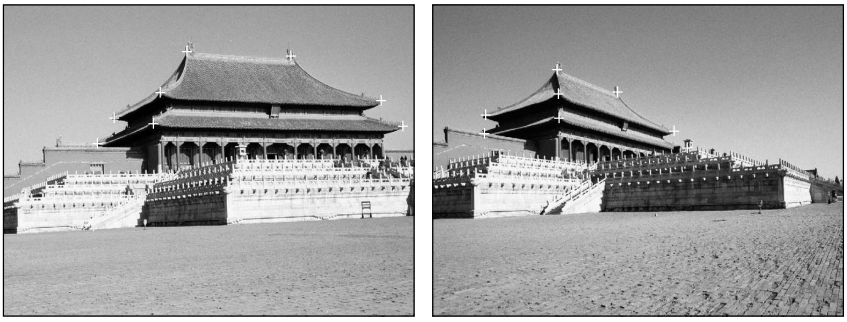

Figure 5.4. Eight pairs of corresponding image points in two views of the Tai-He Palace in the Forbidden City, Beijing, China.

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 must 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 first estimate the null space of χ *ignoring the internal structure of Essential matrix*, obtaining a matrix, say F, which possibly 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.

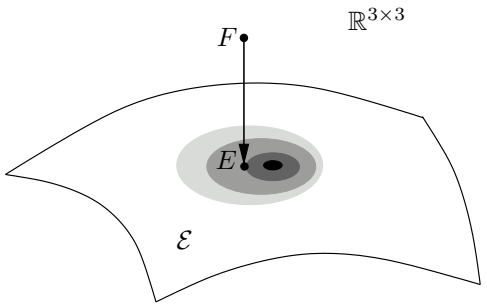

Figure 5.5. Among all points in the Essential space $\mathcal{E} \subset \mathbb{R}^{3\times3}$, E has the shortest Frobenius distance to F. However, the least square error $\|\chi E^s\|^2$ may not be the smallest for E among all points in $\mathcal{E}$. Possible level sets for the value $\|\chi E^s\|^2$ are plotted: the darker the area is, the lower is the value of $\|\chi E^s\|^2$.

Theorem 5.9 (Projection onto the Essential space). *Given a real matrix $F \in \mathbb{R}^{3\times3}$ 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 $E \in \mathcal{E}$ which 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 Frobenious norm (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 $E \in \mathcal{E}_\Sigma$ which 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

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

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 only need to minimize the error function 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 which minimizes this error function is given by $\sigma = (\lambda_1 + \lambda_2)/2$. □

As we have already pointed out, the epipolar constraint only allows for the recovery of the Essential matrix up to a scale (since the epipolar constraint (5.2) is homogeneous in E, it is not modified by multiplying it by any non-zero 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 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 routine does* not *guarantee that the computed U and V are in $SO(3)$. 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 be either operated on $+E$ or $-E$. We leave it as an exercise to the reader that one of the (noisy) matrices $\pm E$ will always has its SVD 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. We leave the details about these four related solutions to the reader as an exercise (see Exercise 5.11).[5]

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 with $\pm E$, 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.

[5] In fact, three of the solutions can be eliminated by imposing the positive depth constraint. See Exercise 5.11.

Algorithm 5.1 (The eight-point algorithm).

For a given set of image correspondences $(\boldsymbol{x}_1^j, \boldsymbol{x}_2^j)$, $j = 1, 2, \ldots, n$ ($n \geq 8$), this algorithm recovers $(R, T) \in SE(3)$ which satisfies

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

1. **Compute a first approximation of the Essential matrix**
   Construct the $\chi = [\boldsymbol{a}^1, \ldots, \boldsymbol{a}^n]^T] \in \mathbb{R}^{n \times 9}$ from correspondences $\boldsymbol{x}_1^j$ and $\boldsymbol{x}_2^j$ as in (5.12), namely
   $$\boldsymbol{a}^j = \boldsymbol{x}_1^j \otimes \boldsymbol{x}_2^j \quad \in \mathbb{R}^9.$$
   Find the vector $E^s \in \mathbb{R}^9$ of unit length such that $\|\chi E^s\|$ is minimized as follows: compute the SVD $\chi = U_\chi \Sigma_\chi V_\chi^T$ and define E^s to be the 9^{th} column of V_χ. Unstack the 9 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 Essential space is $U \Sigma V^T$, where $\Sigma = \text{diag}\{1, 1, 0\}$.

3. **Recover displacement from the Essential matrix**
   We now only need U and V to extract R and T from the Essential matrix as
   $$R = U R_Z^T \left(\pm \frac{\pi}{2}\right) V^T, \quad \widehat{T} = U R_Z \left(\pm \frac{\pi}{2}\right) \Sigma U^T.$$

Example 5.11 (A numerical example). Suppose that

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

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 non-zero 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 direct to compute the four possible decompositions $(R, \widehat{T})$ for E

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

Clearly, the 3rd solution is exactly the original motion $(R, \widehat{T})$ except that the translation T is recovered up to a scale (i.e. it is normalized always to 1). ■

Despite its simplicity, the above algorithm, when used in practice, have a few caveats that are discussed below.

Number of points

The number of 8 points 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 5 degrees of freedom: 3 for rotation and 2 for translation (up to a scale). By utilizing some additional algebraic properties of E, we may reduce the number of points necessary. 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 kernel of χ. Nevertheless, there is usually only one $\alpha \in \mathbb{R}$ such that

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

Therefore, 7 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 6 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 only needs 5 points in general position to recover (R, T). It can be shown that there are up to a total 10 (possibly complex) solutions, though the solutions are not obtainable in closed form.

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 potential 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).

Motion requirement: sufficient baseline

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 epipolar 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 $T \neq 0$. The translation T induces in the image plane so called "parallax". 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 baseline" for the algorithm to be well conditioned. It has been observed experimentally that, even for purely rotational motion $T = 0$, the "spurious" translation created by noise in the image measurements is sufficient for the eight-point algorithm return a correct estimate of R.

Structure requirement: general position

In order for the above algorithm to work properly, the condition that the given 8 points are in "general position" is very important. It can be easily shown that if these points form certain degenerate configurations, the so-called "critical surfaces", the algorithm will fail (see Exercise 5.14). A case of some practical importance is 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).

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 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 either moves according to 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 systematically study the multiple-motion problem in Chapter 7.

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 of linear and angular velocities of the camera.

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 $\gamma = 1$ which is equivalent to choosing the length of translation to be of 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 scales λ's and the motion scales 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, knowing λ_1, 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 $\widehat{\boldsymbol{x}_2}$. It yields

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

This is equivalent to solving the linear equation

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

where $M^j = \left[\widehat{\boldsymbol{x}_2^j} R \boldsymbol{x}_1^j, \ \widehat{\boldsymbol{x}_2^j} T \right] \in \mathbb{R}^{3\times 2}$ and $\bar{\lambda}^j = [\lambda_1^j, \gamma]^T \in \mathbb{R}^2$ $j = 1, 2, \ldots, n$. 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.