12

Structure Computation

This chapter describes how to compute the position of a point in 3-space given its image in two views and the camera matrices of those views. It is assumed that there are errors only in the measured image coordinates, not in the projection matrices $\mathtt{P}$, $\mathtt{P}'$.

Under these circumstances naïve triangulation by back-projecting rays from the measured image points will fail, because the rays will not intersect in general. It is thus necessary to *estimate* a best solution for the point in 3-space.

A best solution requires the definition and minimization of a suitable cost function. This problem is especially critical in affine and projective reconstruction in which there is no meaningful metric information about the object space. It is desirable to find a triangulation method that is invariant to projective transformations of space.

In the following sections we describe the estimation of $\mathbf{X}$ and of its covariance. An optimal (MLE) estimator for the point is developed, and it is shown that a solution can be obtained without requiring numerical minimization.

Note, this is the scenario where $\mathtt{F}$ is given *a priori* and then $\mathbf{X}$ is determined. An alternative scenario is where $\mathtt{F}$ and $\{\mathbf{X}_i\}$ are estimated simultaneously from the image point correspondences $\{\mathbf{x}_i \leftrightarrow \mathbf{x}'_i\}$, but this is not considered in this chapter. It may be solved using the Gold Standard algorithm of section 11.4.1, using the method considered in this chapter as an initial estimate.

12.1 Problem statement

It is supposed that the camera matrices, and hence the fundamental matrix, are provided; or that the fundamental matrix is provided, and hence a pair of consistent camera matrices can be constructed (as in section 9.5(*p*253)). In either case it is assumed that these matrices are known exactly, or at least with great accuracy compared with a pair of matching points in the two images.

Since there are errors in the *measured* points $\mathbf{x}$ and $\mathbf{x}'$, the rays back-projected from the points are skew. This means that there will *not* be a point $\mathbf{X}$ which exactly satisfies $\mathbf{x} = \mathtt{P}\mathbf{X},\ \mathbf{x}' = \mathtt{P}'\mathbf{X}$; and that the image points do *not* satisfy the epipolar constraint $\mathbf{x}'^{\mathsf{T}}\mathtt{F}\mathbf{x} = 0$. These statements are equivalent since the two rays corresponding to a matching pair of points $\mathbf{x} \leftrightarrow \mathbf{x}'$ will meet in space if and only if the points satisfy the epipolar constraint. See figure 12.1.

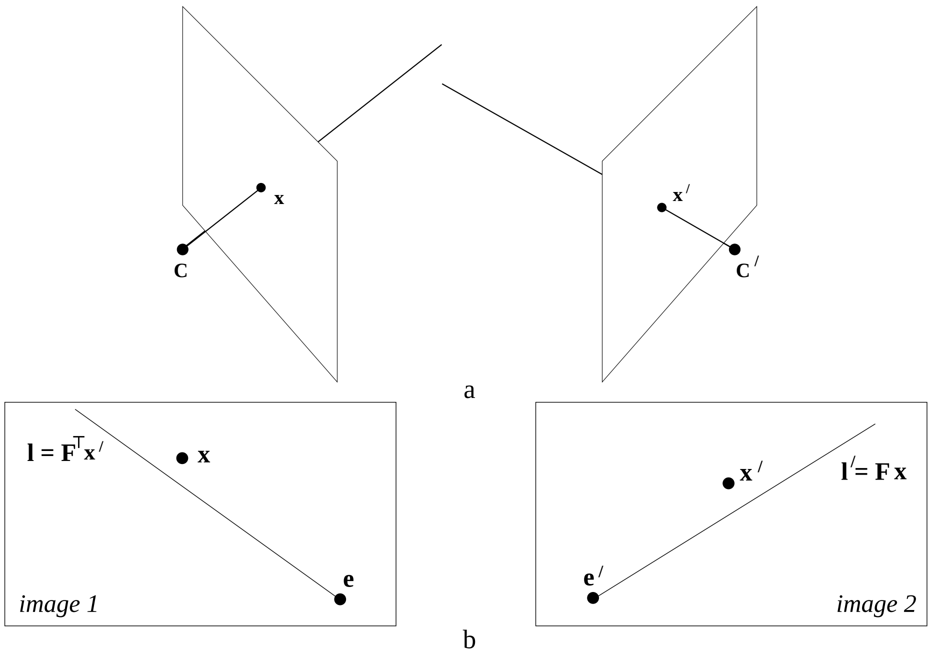

Fig. 12.1. *(a) Rays back-projected from imperfectly measured points* $\mathbf{x}, \mathbf{x}'$ *are skew in 3-space in general. (b) The epipolar geometry for* $\mathbf{x}, \mathbf{x}'$*. The measured points do not satisfy the epipolar constraint. The epipolar line* $\mathbf{l}' = \mathtt{F}\mathbf{x}$ *is the image of the ray through* $\mathbf{x}$*, and* $\mathbf{l} = \mathtt{F}^{\mathsf{T}}\mathbf{x}'$ *is the image of the ray through* $\mathbf{x}'$*. Since the rays do not intersect,* $\mathbf{x}'$ *does not lie on* $\mathbf{l}'$*, and* $\mathbf{x}$ *does not lie on* $\mathbf{l}$.

A desirable feature of the method of triangulation used is that it should be invariant under transformations of the appropriate class for the reconstruction – if the camera matrices are known only up to an affine (or projective) transformation, then it is clearly desirable to use an affine (resp. projective) invariant triangulation method to compute the 3D space points. Thus, denote by τ a triangulation method used to compute a 3D space point $\mathbf{X}$ from a point correspondence $\mathbf{x} \leftrightarrow \mathbf{x}'$ and a pair of camera matrices $\mathtt{P}$ and $\mathtt{P}'$. We write

$$\mathbf{X} = \tau(\mathbf{x}, \mathbf{x}', \mathtt{P}, \mathtt{P}').$$

The triangulation is said to be invariant under a transformation $\mathtt{H}$ if

$$\tau(\mathbf{x}, \mathbf{x}', \mathtt{P}, \mathtt{P}') = \mathtt{H}^{-1}\tau(\mathbf{x}, \mathbf{x}', \mathtt{P}\mathtt{H}^{-1}, \mathtt{P}'\mathtt{H}^{-1}).$$

This means that triangulation using the transformed cameras results in the transformed point.

It is clear, particularly for projective reconstruction, that it is inappropriate to minimize errors in the 3D projective space, $\mathbb{P}^3$. For instance, the method that finds the midpoint of the common perpendicular to the two rays in space is not suitable for projective reconstruction, since concepts such as distance and perpendicularity are not valid in the context of projective geometry. In fact, in projective reconstruction, this method will give different results depending on which particular projective reconstruction is considered – the method is not projective-invariant.

Here we will give a triangulation method that is projective-invariant. The key idea is to estimate a 3D point $\widehat{\mathbf{X}}$ which exactly satisfies the supplied camera geometry, so it projects as

$$\hat{\mathbf{x}} = \mathtt{P}\widehat{\mathbf{X}} \qquad \hat{\mathbf{x}}' = \mathtt{P}'\widehat{\mathbf{X}}$$

and the aim is to estimate $\widehat{\mathbf{X}}$ from the image measurements $\mathbf{x}$ and $\mathbf{x}'$. As described in section 12.3 the maximum likelihood estimate, under Gaussian noise, is given by the point $\widehat{\mathbf{X}}$ which minimizes the reprojection error – the (summed squared) distances between the projections of $\widehat{\mathbf{X}}$ and the measured image points.

Such a triangulation method is projective-invariant because only image distances are minimized, and the points $\hat{\mathbf{x}}$ and $\hat{\mathbf{x}}'$ which are the projections of $\widehat{\mathbf{X}}$ do not depend on the projective frame in which $\widehat{\mathbf{X}}$ is defined, i.e. a different projective reconstruction will project to the same points.

In the following sections simple linear triangulation methods are given. Then the MLE is defined, and it is shown that an optimal solution can be obtained via the root of a sixth-degree polynomial, thus avoiding a non-linear minimization of a cost function.

12.2 Linear triangulation methods

In this section, we describe simple linear triangulation methods. As usual the estimated point does not exactly satisfy the geometric relations, and is not an optimal estimate.

The linear triangulation method is the direct analogue of the DLT method described in section 4.1(*p*88). In each image we have a measurement $\mathbf{x} = \mathtt{P}\mathbf{X}$, $\mathbf{x}' = \mathtt{P}'\mathbf{X}$, and these equations can be combined into a form $\mathtt{A}\mathbf{X} = \mathbf{0}$, which is an equation linear in $\mathbf{X}$.

First the homogeneous scale factor is eliminated by a cross product to give three equations for each image point, of which two are linearly independent. For example for the first image, $\mathbf{x} \times (\mathtt{P}\mathbf{X}) = \mathbf{0}$ and writing this out gives

$$\begin{aligned} x(\mathbf{p}^{3\mathsf{T}}\mathbf{X}) - (\mathbf{p}^{1\mathsf{T}}\mathbf{X}) &= 0 \\ y(\mathbf{p}^{3\mathsf{T}}\mathbf{X}) - (\mathbf{p}^{2\mathsf{T}}\mathbf{X}) &= 0 \\ x(\mathbf{p}^{2\mathsf{T}}\mathbf{X}) - y(\mathbf{p}^{1\mathsf{T}}\mathbf{X}) &= 0 \end{aligned}$$

where $\mathbf{p}^{i\mathsf{T}}$ are the rows of $\mathtt{P}$. These equations are *linear* in the components of $\mathbf{X}$.

An equation of the form $\mathtt{A}\mathbf{X} = \mathbf{0}$ can then be composed, with

$$\mathtt{A} = \begin{bmatrix} x\mathbf{p}^{3\mathsf{T}} - \mathbf{p}^{1\mathsf{T}} \\ y\mathbf{p}^{3\mathsf{T}} - \mathbf{p}^{2\mathsf{T}} \\ x'\mathbf{p}'^{3\mathsf{T}} - \mathbf{p}'^{1\mathsf{T}} \\ y'\mathbf{p}'^{3\mathsf{T}} - \mathbf{p}'^{2\mathsf{T}} \end{bmatrix}$$

where two equations have been included from each image, giving a total of four equations in four homogeneous unknowns. This is a redundant set of equations, since the solution is determined only up to scale. Two ways of solving the set of equations of the form $\mathtt{A}\mathbf{X} = \mathbf{0}$ were discussed in section 4.1(*p*88) and will be considered again here.

Homogeneous method (DLT). The method of section 4.1.1(*p*90) finds the solution as the unit singular vector corresponding to the smallest singular value of $\mathtt{A}$, as shown in section A5.3(*p*592). The discussion in section 4.1.1 on the merits of normalization, and of including two or three equations from each image, applies equally well here.

Inhomogeneous method. In section 4.1.2(*p*90) the solution of this system as a set of inhomogeneous equations is discussed. By setting $\mathbf{X} = (\mathrm{X}, \mathrm{Y}, \mathrm{Z}, 1)^\mathsf{T}$ the set of homogeneous equations, $\mathtt{A}\mathbf{X} = \mathbf{0}$, is reduced to a set of four inhomogeneous equations in three unknowns. The least-squares solution to these inhomogeneous equations is described in section A5.1(*p*588). As explained in section 4.1.2, however, difficulties arise if the true solution $\mathbf{X}$ has last coordinate equal or close to 0. In this case, it is not legitimate to set it to 1 and instabilities can occur.

Discussion. These two methods are quite similar, but in fact have quite different properties in the presence of noise. The inhomogeneous method assumes that the solution point $\mathbf{X}$ is not at infinity, for otherwise we could not assume that $\mathbf{X} = (x, y, z, 1)^\mathsf{T}$. This is a disadvantage of this method when we are seeking to carry out a projective reconstruction, where reconstructed points may lie on the plane at infinity. Furthermore, neither of these two linear methods is quite suitable for projective reconstruction, since they are not projective-invariant. To see this, suppose that camera matrices $\mathtt{P}$ and $\mathtt{P}'$ are replaced by $\mathtt{PH}^{-1}$ and $\mathtt{P}'\mathtt{H}^{-1}$. One sees that in this case the matrix of equations, $\mathtt{A}$, becomes $\mathtt{AH}^{-1}$. A point $\mathbf{X}$ such that $\mathtt{A}\mathbf{X} = \boldsymbol{\epsilon}$ for the original problem corresponds to a point $\mathtt{H}\mathbf{X}$ satisfying $(\mathtt{AH}^{-1})(\mathtt{H}\mathbf{X}) = \boldsymbol{\epsilon}$ for the transformed problem. Thus, there is a one-to-one correspondence between points $\mathbf{X}$ and $\mathtt{H}\mathbf{X}$ giving the same error. However, neither the condition $\|\mathbf{X}\| = 1$ for the homogeneous method, nor the condition $\mathbf{X} = (\mathrm{X}, \mathrm{Y}, \mathrm{Z}, 1)^\mathsf{T}$ for the inhomogeneous method, is invariant under application of the projective transformation $\mathtt{H}$. Thus, in general the point $\mathbf{X}$ solving the original problem will not correspond to a solution $\mathtt{H}\mathbf{X}$ for the transformed problem.

For affine transformations, on the other hand, the situation is different. In fact, although the condition $\|\mathbf{X}\| = 1$ is not preserved under affine transformations, the condition $\mathbf{X} = (\mathrm{X}, \mathrm{Y}, \mathrm{Z}, 1)^\mathsf{T}$ is preserved, since for an affine transformation, $\mathtt{H}(\mathrm{X}, \mathrm{Y}, \mathrm{Z}, 1)^\mathsf{T} = (\mathrm{X}', \mathrm{Y}', \mathrm{Z}', 1)^\mathsf{T}$. This means that there is a one-to-one correspondence between a vector $\mathbf{X} = (\mathrm{X}, \mathrm{Y}, \mathrm{Z}, 1)^\mathsf{T}$ such that $\mathtt{A}(x, y, z, 1)^\mathsf{T} = \boldsymbol{\epsilon}$ and the vector $\mathtt{H}\mathbf{X} = (\mathrm{X}', \mathrm{Y}', \mathrm{Z}', 1)^\mathsf{T}$ such that $(\mathtt{AH}^{-1})(\mathrm{X}', \mathrm{Y}', \mathrm{Z}', 1)^\mathsf{T} = \boldsymbol{\epsilon}$. The error is the same for corresponding points. Thus, the points that minimize the error $\|\boldsymbol{\epsilon}\|$ correspond as well. Hence, the inhomogeneous method is affine-invariant, whereas the homogeneous method is not.

In the remainder of this chapter we will describe a method for triangulation that is invariant to the projective frame of the cameras, and minimizes a geometric image error. This will be the recommended triangulation method. Nevertheless, the homogeneous linear method described above often provides acceptable results. Furthermore, it has the virtue that it generalizes easily to triangulation when more than two views of the point are available.

12.3 Geometric error cost function

A typical observation consists of a noisy point correspondence $\mathbf{x} \leftrightarrow \mathbf{x}'$ which does not in general satisfy the epipolar constraint. In reality, the correct values of the cor-

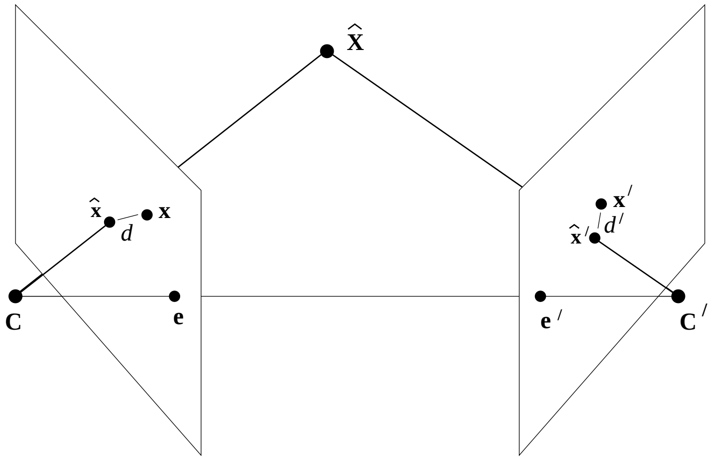

Fig. 12.2. **Minimization of geometric error.** *The estimated 3-space point $\widehat{\mathbf{X}}$ projects to the two images at $\hat{\mathbf{x}}, \hat{\mathbf{x}}'$. The corresponding image points $\hat{\mathbf{x}}, \hat{\mathbf{x}}'$ satisfy the epipolar constraint, unlike the measured points $\mathbf{x}$ and $\mathbf{x}'$. The point $\widehat{\mathbf{X}}$ is chosen so that the reprojection error $d^2 + d'^2$ is minimized.*

responding image points should be points $\bar{\mathbf{x}} \leftrightarrow \bar{\mathbf{x}}'$ lying close to the measured points $\mathbf{x} \leftrightarrow \mathbf{x}'$ and satisfying the epipolar constraint $\bar{\mathbf{x}}'^{\mathsf{T}}\mathtt{F}\bar{\mathbf{x}} = 0$ exactly.

We seek the points $\hat{\mathbf{x}}$ and $\hat{\mathbf{x}}'$ that minimize the function

$$\mathcal{C}(\mathbf{x}, \mathbf{x}') = d(\mathbf{x}, \hat{\mathbf{x}})^2 + d(\mathbf{x}', \hat{\mathbf{x}}')^2 \quad \text{subject to} \quad \hat{\mathbf{x}}'^{\mathsf{T}}\mathtt{F}\hat{\mathbf{x}} = 0 \tag{12.1}$$

where $d(*, *)$ is the Euclidean distance between the points. This is equivalent to minimizing the reprojection error for a point $\widehat{\mathbf{X}}$ which is mapped to $\hat{\mathbf{x}}$ and $\hat{\mathbf{x}}'$ by projection matrices consistent with $\mathtt{F}$, as illustrated in figure 12.2.

As explained in section 4.3(*p*102), assuming a Gaussian error distribution, the points $\hat{\mathbf{x}}'$ and $\hat{\mathbf{x}}$ are Maximum Likelihood Estimates (MLE) for the true image point correspondences. Once $\hat{\mathbf{x}}'$ and $\hat{\mathbf{x}}$ are found, the point $\widehat{\mathbf{X}}$ may be found by any triangulation method, since the corresponding rays will meet precisely in space.

This cost function could, of course, be minimized using a numerical minimization method such as Levenberg–Marquardt (section A6.2(*p*600)). A close approximation to the minimum may also be found using a first-order approximation to the geometric cost function, namely the Sampson error, as described in the next section. However, in section 12.5 it is shown that the minimum can be obtained non-iteratively by the solution of a sixth-degree polynomial.

12.4 Sampson approximation (first-order geometric correction)

Before deriving the exact polynomial solution we develop the Sampson approximation, which is valid when the measurement errors are small compared with the measurements. The Sampson approximation to the geometric *cost* function in the case of the fundamental matrix has already been discussed in section 11.4.3. Here we are concerned with computing the *correction* to the measured points.

The Sampson correction $\boldsymbol{\delta}_{\mathbf{X}}$ to the measured point $\mathbf{X} = (x, y, x', y')^{\mathsf{T}}$ (note, in this section $\mathbf{X}$ does not denote a homogeneous 3-space point) is shown in section 4.2.6(*p*98)