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.

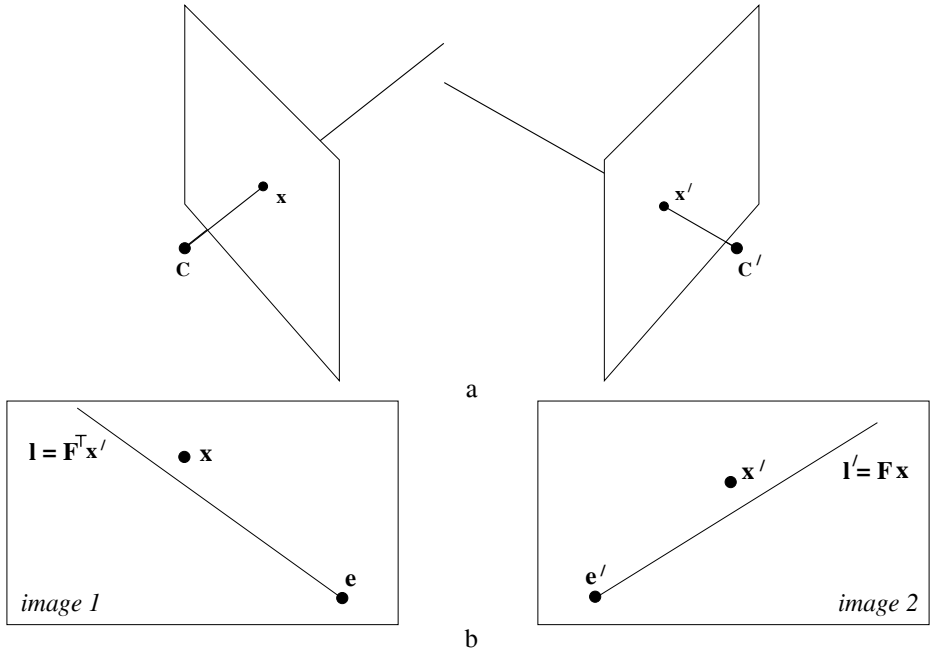

Fig. 12.1. *(a) Rays back-projected from imperfectly measured points* $\mathbf{x}, \mathbf{x}'$ *are skew in 3-space in general. (b) The epipolar geometry for* $\mathbf{x}, \mathbf{x}'$*. The measured points do not satisfy the epipolar constraint. The epipolar line* $\mathbf{l}' = \mathtt{F}\mathbf{x}$ *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

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

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 $\mathbf{X} = (\mathtt{X}, \mathtt{Y}, \mathtt{Z}, 1)^\mathsf{T}$ 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 $\mathtt{P}\mathtt{H}^{-1}$ and $\mathtt{P}'\mathtt{H}^{-1}$. One sees that in this case the matrix of equations, $\mathtt{A}$, becomes $\mathtt{A}\mathtt{H}^{-1}$. 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 $(\mathtt{A}\mathtt{H}^{-1})(\mathtt{H}\mathbf{X}) = \boldsymbol{\epsilon}$ 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 $\mathbf{X} = (\mathtt{X}, \mathtt{Y}, \mathtt{Z}, 1)^\mathsf{T}$ 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 $\mathbf{X} = (\mathtt{X}, \mathtt{Y}, \mathtt{Z}, 1)^\mathsf{T}$ is preserved, since for an affine transformation, $\mathtt{H}(\mathtt{X}, \mathtt{Y}, \mathtt{Z}, 1)^\mathsf{T} = (\mathtt{X}', \mathtt{Y}', \mathtt{Z}', 1)^\mathsf{T}$. This means that there is a one-to-one correspondence between a vector $\mathbf{X} = (\mathtt{X}, \mathtt{Y}, \mathtt{Z}, 1)^\mathsf{T}$ such that $\mathtt{A}(x, y, z, 1)^\mathsf{T} = \boldsymbol{\epsilon}$ and the vector $\mathtt{H}\mathbf{X} = (\mathtt{X}', \mathtt{Y}', \mathtt{Z}', 1)^\mathsf{T}$ such that $(\mathtt{A}\mathtt{H}^{-1})(\mathtt{X}', \mathtt{Y}', \mathtt{Z}', 1)^\mathsf{T} = \boldsymbol{\epsilon}$. 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-

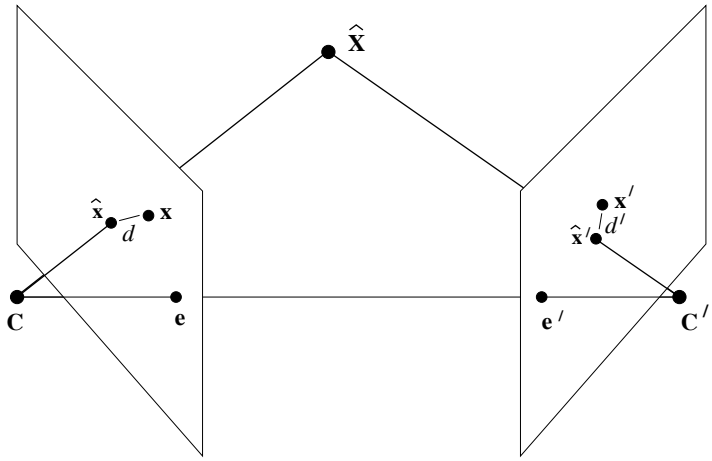

Fig. 12.2. **Minimization of geometric error.** *The estimated 3-space point $\widehat{\mathbf{X}}$ projects to the two images at $\hat{\mathbf{x}}, \hat{\mathbf{x}}'$. The corresponding image points $\hat{\mathbf{x}}, \hat{\mathbf{x}}'$ satisfy the epipolar constraint, unlike the measured points $\mathbf{x}$ 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)

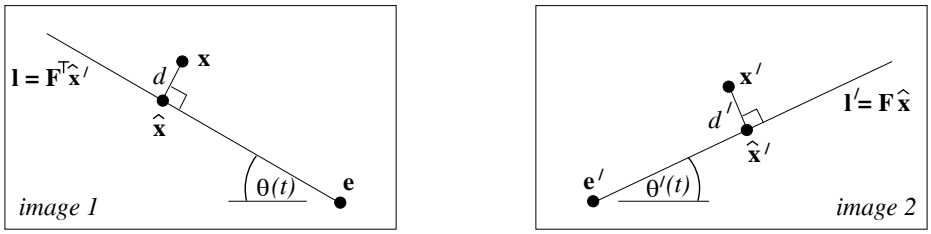

Fig. 12.3. *The projections $\hat{\mathbf{x}}$ and $\hat{\mathbf{x}}'$ of an estimated 3D point $\widehat{\mathbf{X}}$ lie on a pair of corresponding epipolar lines. The optimal $\hat{\mathbf{x}}$ and $\hat{\mathbf{x}}'$ will lie at the foot of the perpendiculars from the measured points $\mathbf{x}$ and $\mathbf{x}'$. Parametrizing the corresponding epipolar lines as a one-parameter family, the optimal estimation of $\widehat{\mathbf{X}}$ is reduced to a one-parameter search for corresponding epipolar lines so as to minimize the squared sum of perpendicular distances $d^2 + d'^2$.*

to be (4.11–*p*99)

$$\boldsymbol{\delta}_{\mathbf{X}} = -\mathtt{J}^{\mathsf{T}}(\mathtt{J}\mathtt{J}^{\mathsf{T}})^{-1}\boldsymbol{\epsilon}$$

and the corrected point is

$$\widehat{\mathbf{X}} = \mathbf{X} + \boldsymbol{\delta}_{\mathbf{X}} = \mathbf{X} - \mathtt{J}^{\mathsf{T}}(\mathtt{J}\mathtt{J}^{\mathsf{T}})^{-1}\boldsymbol{\epsilon}.$$

As shown in section 11.4.3 in the case of the variety defined by $\mathbf{x}'^{\mathsf{T}}\mathtt{F}\mathbf{x} = 0$, the error $\boldsymbol{\epsilon} = \mathbf{x}'^{\mathsf{T}}\mathtt{F}\mathbf{x}$, and the Jacobian is

$$J = \partial\boldsymbol{\epsilon}/\partial x = [(\mathtt{F}^{\mathsf{T}}\mathbf{x}')_1, (\mathtt{F}^{\mathsf{T}}\mathbf{x}')_2, (\mathtt{F}\mathbf{x})_1, (\mathtt{F}\mathbf{x})_2]$$

where for instance $(\mathtt{F}^{\mathsf{T}}\mathbf{x}')_1 = f_{11}x' + f_{21}y' + f_{31}$, etc. Then the first-order approximation to the corrected point is simply

$$\begin{pmatrix} \hat{x} \\ \hat{y} \\ \hat{x}' \\ \hat{y}' \end{pmatrix} = \begin{pmatrix} x \\ y \\ x' \\ y' \end{pmatrix} - \frac{\mathbf{x}'^{\mathsf{T}}\mathtt{F}\mathbf{x}}{(\mathtt{F}\mathbf{x})_1^2 + (\mathtt{F}\mathbf{x})_2^2 + (\mathtt{F}^{\mathsf{T}}\mathbf{x}')_1^2 + (\mathtt{F}^{\mathsf{T}}\mathbf{x}')_2^2} \begin{pmatrix} (\mathtt{F}^{\mathsf{T}}\mathbf{x}')_1 \\ (\mathtt{F}^{\mathsf{T}}\mathbf{x}')_2 \\ (\mathtt{F}\mathbf{x})_1 \\ (\mathtt{F}\mathbf{x})_2 \end{pmatrix}.$$

The approximation is accurate if the correction in each image is small (less than a pixel), and is cheap to compute. Note, however, that the corrected points will *not* satisfy the epipolar relation $\hat{\mathbf{x}}'^{\mathsf{T}}\mathtt{F}\hat{\mathbf{x}} = 0$ exactly. The method of the following section computes the points $\hat{\mathbf{x}}, \hat{\mathbf{x}}'$ which do exactly satisfy the epipolar constraint, but is more costly.

12.5 An optimal solution

In this section, we describe a method of triangulation that finds the global minimum of the cost function (12.1) using a non-iterative algorithm. If the Gaussian noise model can be assumed to be correct, this triangulation method is then provably optimal.

12.5.1 Reformulation of the minimization problem

Given a measured correspondence $\mathbf{x} \leftrightarrow \mathbf{x}'$, we seek a pair of points $\hat{\mathbf{x}}'$ and $\hat{\mathbf{x}}$ that minimize the sum of squared distances (12.1) subject to the epipolar constraint $\hat{\mathbf{x}}'^{\mathsf{T}}\mathtt{F}\hat{\mathbf{x}} = 0$.

The following discussion relates to figure 12.3. Any pair of points satisfying the

epipolar constraint must lie on a pair of corresponding epipolar lines in the two images. Thus, in particular, the optimum point $\hat{\mathbf{x}}$ lies on an epipolar line $\mathbf{l}$ and $\hat{\mathbf{x}}'$ lies on the corresponding epipolar line $\mathbf{l}'$. On the other hand, any other pair of points lying on the lines $\mathbf{l}$ and $\mathbf{l}'$ will also satisfy the epipolar constraint. This is true in particular for the point $\mathbf{x}_\perp$ on $\mathbf{l}$ lying closest to the measured point $\mathbf{x}$, and the correspondingly defined point $\mathbf{x}'_\perp$ on $\mathbf{l}'$. Of all pairs of points on the lines $\mathbf{l}$ and $\mathbf{l}'$, the points $\mathbf{x}_\perp$ and $\mathbf{x}'_\perp$ minimize the squared distance sum of (12.1). It follows that $\hat{\mathbf{x}}' = \mathbf{x}'_\perp$ and $\hat{\mathbf{x}} = \mathbf{x}_\perp$, where $\mathbf{x}_\perp$ and $\mathbf{x}'_\perp$ are defined with respect to a pair of matching epipolar lines $\mathbf{l}$ and $\mathbf{l}'$. Consequently, we may write $d(\mathbf{x}, \hat{\mathbf{x}}) = d(\mathbf{x}, \mathbf{l})$, where $d(\mathbf{x}, \mathbf{l})$ represents the perpendicular distance from the point $\mathbf{x}$ to the line $\mathbf{l}$. A similar expression holds for $d(\mathbf{x}', \hat{\mathbf{x}}')$.

In view of the previous paragraph, we may formulate the minimization problem differently as follows. We seek to minimize

$$d(\mathbf{x}, \mathbf{l})^2 + d(\mathbf{x}', \mathbf{l}')^2 \tag{12.2}$$

where $\mathbf{l}$ and $\mathbf{l}'$ range over all choices of corresponding epipolar lines. The point $\hat{\mathbf{x}}$ is then the closest point on the line $\mathbf{l}$ to the point $\mathbf{x}$ and the point $\hat{\mathbf{x}}'$ is similarly defined.

Our strategy for minimizing (12.2) is as follows:

(i) Parametrize the pencil of epipolar lines in the first image by a parameter t. Thus an epipolar line in the first image may be written as $\mathbf{l}(t)$.
(ii) Using the fundamental matrix $\mathtt{F}$, compute the corresponding epipolar line $\mathbf{l}'(t)$ in the second image.
(iii) Express the distance function $d(\mathbf{x}, \mathbf{l}(t))^2 + d(\mathbf{x}', \mathbf{l}'(t))^2$ explicitly as a function of t.
(iv) Find the value of t that minimizes this function.

In this way, the problem is reduced to that of finding the minimum of a function of a single variable t, i.e.

$$\min_{\hat{\mathbf{x}}} \mathcal{C} = d(\mathbf{x}, \hat{\mathbf{x}})^2 + d(\mathbf{x}', \hat{\mathbf{x}}')^2 = \min_t \mathcal{C} = d(\mathbf{x}, \mathbf{l}(t))^2 + d(\mathbf{x}', \mathbf{l}'(t))^2.$$

It will be seen that for a suitable parametrization of the pencil of epipolar lines the distance function is a rational polynomial function of t. Using techniques of elementary calculus, the minimization problem reduces to finding the real roots of a polynomial of degree 6.

12.5.2 Details of the minimization

If both of the image points correspond with the epipoles, then the point in space lies on the line joining the camera centres. In this case it is impossible to determine the position of the point in space. If only one of the corresponding points lies at an epipole, then we conclude that the point in space must coincide with the other camera centre. Consequently, we assume that neither of the two image points $\mathbf{x}$ and $\mathbf{x}'$ corresponds with an epipole.

In this case, we may simplify the analysis by applying a rigid transformation to each image in order to place both points $\mathbf{x}$ and $\mathbf{x}'$ at the origin, $(0, 0, 1)^\mathsf{T}$ in homogeneous

coordinates. Furthermore, the epipoles may be placed on the x-axis at points $(1, 0, f)^\mathsf{T}$ and $(1, 0, f')^\mathsf{T}$ respectively. A value f equal to 0 means that the epipole is at infinity. Applying these two rigid transforms has no effect on the sum-of-squares distance function in (12.1), and hence does not change the minimization problem.

Thus, in future we assume that in homogeneous coordinates, $\mathbf{x} = \mathbf{x}' = (0, 0, 1)^\mathsf{T}$ and that the two epipoles are at points $(1, 0, f)^\mathsf{T}$ and $(1, 0, f')^\mathsf{T}$. In this case, since $\mathtt{F}(1, 0, f)^\mathsf{T} = (1, 0, f')\mathtt{F} = \mathbf{0}$, the fundamental matrix has a special form

$$\mathtt{F} = \begin{pmatrix} ff'd & -f'c & -f'd \\ -fb & a & b \\ -fd & c & d \end{pmatrix}. \tag{12.3}$$

Consider an epipolar line in the first image passing through the point $(0, t, 1)^\mathsf{T}$ (still in homogeneous coordinates) and the epipole $(1, 0, f)^\mathsf{T}$. We denote this epipolar line by $\mathbf{l}(t)$. The vector representing this line is given by the cross product $(0, t, 1) \times (1, 0, f) = (tf, 1, -t)$, so the squared distance from the line to the origin is

$$d(\mathbf{x}, \mathbf{l}(t))^2 = \frac{t^2}{1 + (tf)^2}.$$

Using the fundamental matrix to find the corresponding epipolar line in the other image, we see that

$$\mathbf{l}'(t) = \mathtt{F}(0, t, 1)^\mathsf{T} = (-f'(ct + d), at + b, ct + d)^\mathsf{T}. \tag{12.4}$$

This is the representation of the line $\mathbf{l}'(t)$ as a homogeneous vector. The squared distance of this line from the origin is equal to

$$d(\mathbf{x}', \mathbf{l}'(t))^2 = \frac{(ct + d)^2}{(at + b)^2 + f'^2(ct + d)^2}.$$

The total squared distance is therefore given by

$$s(t) = \frac{t^2}{1 + f^2t^2} + \frac{(ct + d)^2}{(at + b)^2 + f'^2(ct + d)^2}. \tag{12.5}$$

Our task is to find the minimum of this function.

We may find the minimum using techniques of elementary calculus, as follows. We compute the derivative

$$s'(t) = \frac{2t}{(1 + f^2t^2)^2} - \frac{2(ad - bc)(at + b)(ct + d)}{((at + b)^2 + f'^2(ct + d)^2)^2}. \tag{12.6}$$

Maxima and minima of $s(t)$ will occur when $s'(t) = 0$. Collecting the two terms in $s'(t)$ over a common denominator and equating the numerator to 0 gives a condition

$$\begin{aligned} g(t) &= t((at + b)^2 + f'^2(ct + d)^2)^2 \\ &\quad -(ad - bc)(1 + f^2t^2)^2(at + b)(ct + d) \\ &= 0. \end{aligned} \tag{12.7}$$

The minima and maxima of $s(t)$ will occur at the roots of this polynomial. This is a

Objective

Given a measured point correspondence $\mathbf{x} \leftrightarrow \mathbf{x}'$, and a fundamental matrix $\mathtt{F}$, compute the corrected correspondences $\hat{\mathbf{x}} \leftrightarrow \hat{\mathbf{x}}'$ that minimize the geometric error (12.1) subject to the epipolar constraint $\hat{\mathbf{x}}'^{\mathsf{T}}\mathtt{F}\hat{\mathbf{x}} = 0$.

Algorithm

(i) Define transformation matrices

$$\mathtt{T} = \begin{bmatrix} 1 & & -x \\ & 1 & -y \\ & & 1 \end{bmatrix} \text{ and } \mathtt{T}' = \begin{bmatrix} 1 & & -x' \\ & 1 & -y' \\ & & 1 \end{bmatrix}.$$

These are the translations that take $\mathbf{x} = (x, y, 1)^{\mathsf{T}}$ and $\mathbf{x}' = (x', y', 1)^{\mathsf{T}}$ to the origin.

(ii) Replace $\mathtt{F}$ by $\mathtt{T}'^{-\mathsf{T}}\mathtt{F}\mathtt{T}^{-1}$. The new $\mathtt{F}$ corresponds to translated coordinates.

(iii) Compute the right and left epipoles $\mathbf{e} = (e_1, e_2, e_3)^{\mathsf{T}}$ and $\mathbf{e}' = (e'_1, e'_2, e'_3)^{\mathsf{T}}$ such that $\mathbf{e}'^{\mathsf{T}}\mathtt{F} = \mathbf{0}$ and $\mathtt{F}\mathbf{e} = \mathbf{0}$. Normalize (multiply by a scale) $\mathbf{e}$ such that $e_1^2 + e_2^2 = 1$ and do the same to $\mathbf{e}'$.

(iv) Form matrices

$$\mathtt{R} = \begin{bmatrix} e_1 & e_2 & \\ -e_2 & e_1 & \\ & & 1 \end{bmatrix} \text{ and } \mathtt{R}' = \begin{bmatrix} e'_1 & e'_2 & \\ -e'_2 & e'_1 & \\ & & 1 \end{bmatrix}$$

and observe that $\mathtt{R}$ and $\mathtt{R}'$ are rotation matrices, and $\mathtt{R}\mathbf{e} = (1, 0, e_3)^{\mathsf{T}}$ and $\mathtt{R}'\mathbf{e}' = (1, 0, e'_3)^{\mathsf{T}}$.

(v) Replace $\mathtt{F}$ by $\mathtt{R}'\mathtt{F}\mathtt{R}^{\mathsf{T}}$. The resulting $\mathtt{F}$ must have the form (12.3).

(vi) Set $f = e_3$, $f' = e'_3$, $a = F_{22}$, $b = F_{23}$, $c = F_{32}$ and $d = F_{33}$.

(vii) Form the polynomial $g(t)$ as a polynomial in t according to (12.7). Solve for t to get 6 roots.

(viii) Evaluate the cost function (12.5) at the real part of each of the roots of $g(t)$ (alternatively evaluate at only the real roots of $g(t)$). Also, find the asymptotic value of (12.1) for $t = \infty$, namely $1/f^2 + c^2/(a^2 + f'^2c^2)$. Select the value t_{min} of t that gives the smallest value of the cost function.

(ix) Evaluate the two lines $\mathbf{l} = (tf, 1, -t)$ and $\mathbf{l}'$ given by (12.4) at t_{min} and find $\hat{\mathbf{x}}$ and $\hat{\mathbf{x}}'$ as the closest points on these lines to the origin. For a general line (λ, μ, ν), the formula for the closest point on the line to the origin is $(-\lambda\nu, -\mu\nu, \lambda^2 + \mu^2)$.

(x) Transfer back to the original coordinates by replacing $\hat{\mathbf{x}}$ by $\mathtt{T}^{-1}\mathtt{R}^{\mathsf{T}}\hat{\mathbf{x}}$ and $\hat{\mathbf{x}}'$ by $\mathtt{T}'^{-1}\mathtt{R}'^{\mathsf{T}}\hat{\mathbf{x}}'$.

(xi) The 3-space point $\widehat{\mathbf{X}}$ may then be obtained by the homogeneous method of section 12.2.

Algorithm 12.1. *The optimal triangulation method.*

polynomial of degree 6, which may have up to 6 real roots, corresponding to 3 minima and 3 maxima of the function $s(t)$. The absolute minimum of the function $s(t)$ may be found by finding the roots of $g(t)$ and evaluating the function $s(t)$ given by (12.5) at each of the real roots. More simply, one checks the value of $s(t)$ at the real part of each root (complex or real) of $g(t)$, which saves the trouble of determining if a root is real or complex. One should also check the asymptotic value of $s(t)$ as $t \to \infty$ to see if the minimum distance occurs when $t = \infty$, corresponding to an epipolar line $fx = 1$ in the first image.

The overall method is summarized in algorithm 12.1.

a b

Fig. 12.4. *(a) Example of a cost function with three minima. (b) This is the cost function for a perfect point match, which nevertheless has two minima.*

12.5.3 Local minima

The fact that $g(t)$ in (12.7) has degree 6 means that $s(t)$ may have as many as three minima. In fact, this is indeed possible, as the following case shows. Setting $f = f' = 1$ and

$$\mathtt{F} = \begin{pmatrix} 4 & -3 & -4 \\ -3 & 2 & 3 \\ -4 & 3 & 4 \end{pmatrix}$$

gives a function

$$s(t) = \frac{t^2}{1+t^2} + \frac{(3t+4)^2}{(2t+3)^2 + (3t+4)^2}$$

with graph as shown in figure 12.4a[1]. The three minima are clearly shown.

As a second example, we consider the case where $f = f' = 1$, and

$$\mathtt{F} = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 2 & -1 \\ 0 & 1 & 0 \end{pmatrix}.$$

In this case, the function $s(t)$ is given by

$$s(t) = \frac{t^2}{t^2+1} + \frac{t^2}{t^2 + (2t-1)^2}$$

and both terms of the cost function vanish for a value of $t = 0$, which means that the corresponding points $\mathbf{x}$ and $\mathbf{x}'$ exactly satisfy the epipolar constraint. This can be verified by observing that $\mathbf{x}'^{\mathsf{T}}\mathtt{F}\mathbf{x} = 0$. Thus the two points are exactly matched. A graph of the cost function $s(t)$ is shown in figure 12.4b. Apart from the absolute minimum at $t = 0$ there is also a local minimum at $t = 1$. Thus, even in the case of perfect matches local minima may occur. This example shows that an algorithm that attempts to minimize the cost function in (12.1), or equivalently (12.2), by an iterative

[1] In this graph and also figure 12.4b we make the substitution $t = \tan(\theta)$ and plot for θ in the range $-\pi/2 \leq \theta \leq \pi/2$, so as to show the whole infinite range for t.

search beginning from an arbitrary initial point is in danger of finding a local minimum, even in the case of perfect point matches.

12.5.4 Evaluation on real images

An experiment was carried out using the calibration cube images shown in figure 11.2-(*p*289) with the goal of determining how the triangulation method effects the accuracy of reconstruction. A Euclidean model of the cubes, to be used as ground truth, was estimated and refined using accurate image measurements. The measured pixel locations were corrected to correspond exactly to the Euclidean model, requiring coordinate corrections averaging 0.02 pixels.

At this stage we had a model and a set of matched points corresponding exactly to the model. Next, a projective reconstruction of the points was computed and a projective transformation `H` computed that brought the projective reconstruction into agreement with the Euclidean model. Controlled zero-mean Gaussian noise was introduced into the point coordinates, and triangulation using two methods was carried out in the projective frame, the transformation `H` applied, and the error of each method was measured in the Euclidean frame. Figure 12.5 shows the results of this experiment for the two triangulation methods. The graph shows the average reconstruction error over all points in 10 separate runs at each chosen noise level. It clearly shows that the optimal method gives superior reconstruction results.

In this pair of images the two epipoles are distant from the image. For cases where the epipoles are close to the images, results on synthetic images show that the advantage of the polynomial method will be more pronounced.

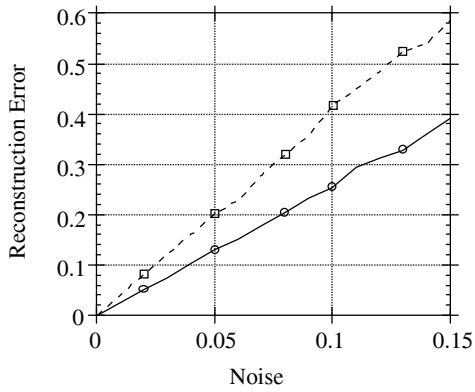

Fig. 12.5. **Reconstruction error comparison of triangulation methods.** *The graph shows the reconstruction error obtained using two triangulation methods: (i) selection of the midpoint of the common perpendicular to the rays in the projective frame (top curve), and (ii) the optimal polynomial method (lower curve). On the horizontal axis is the noise, on the vertical axis the reconstruction error. The units for reconstruction error are relative to a unit distance equal to the side of one of the dark squares in the calibration cube image figure 11.2(*p289*). Even for the best method the error is large for higher noise levels, because there is little movement between the images.*

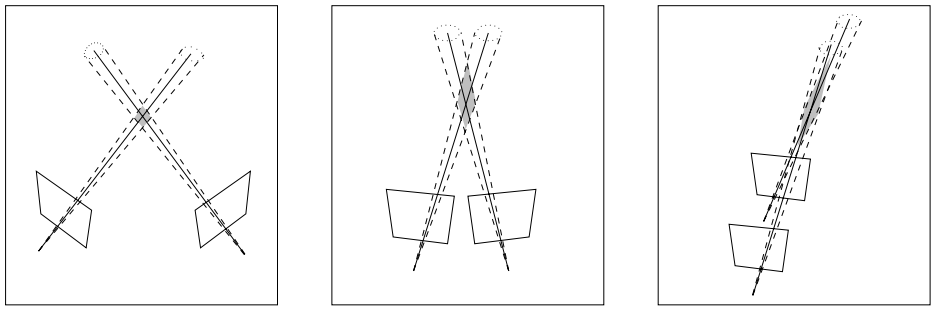

Fig. 12.6. **Uncertainty of reconstruction.** *The shaded region in each case illustrates the shape of the uncertainty region, which depends on the angle between the rays. Points are less precisely localized along the ray as the rays become more parallel. Forward motion in particular can give poor reconstructions since rays are almost parallel for much of the field of view.*

12.6 Probability distribution of the estimated 3D point

An illustration of the distribution of the reconstructed point is given in figure 12.6. A good rule of thumb is that the angle between the rays determines the accuracy of reconstruction. This is a better guide than simply considering the baseline, which is the more commonly used measure.

More formally the probability of a particular 3D point $\mathbf{X}$ depends on the probability of obtaining its image in each view. We will consider a simplified example where the objective is to estimate the probability that a point on a plane has position $\mathbf{X} = (\mathrm{X}, \mathrm{Y})^\mathsf{T}$ given its images $x = f(\mathbf{X})$ and $x' = f'(\mathbf{X})$ in two line cameras. (The projections f and f' are expressible in terms of 2×3 projection matrices $\mathtt{P}_{2\times3}$ and $\mathtt{P}'_{2\times3}$ respectively – see section 6.4.2(*p*175)). The imaging geometry is shown in figure 12.7(a).

Suppose that the measured image point is at x in the first image, and the measurement process is corrupted by Gaussian noise with mean zero and variance σ^2, then the probability of obtaining x, given that the true image point is $f(\mathbf{X})$, is given by

$$p(x|\mathbf{X}) = (2\pi\sigma^2)^{-1/2} \exp\left(-|f(\mathbf{X}) - x|^2/(2\sigma^2)\right).$$

with a similar expression for $p(x'|\mathbf{X})$. We wish to compute the *a posteriori* distribution:

$$p(\mathbf{X}|x, x') = p(x, x'|\mathbf{X})p(\mathbf{X})/p(x, x').$$

Assuming a uniform prior probability $p(\mathbf{X})$, and independent image measurements in the two images, it follows that

$$p(\mathbf{X}|x, x') \sim p(x, x'|\mathbf{X}) = p(x|\mathbf{X})p(x'|\mathbf{X}).$$

Figure 12.7 shows an example of this Probability Density Function (PDF). The bias and variance of this example is discussed in appendix 3(*p*568).

12.7 Line reconstruction

Suppose a line in 3-space is projected to lines in two views as $\mathbf{l}$ and $\mathbf{l}'$. The line in 3-space can be reconstructed by back-projecting each line to give a plane in 3-space, and intersecting the planes.

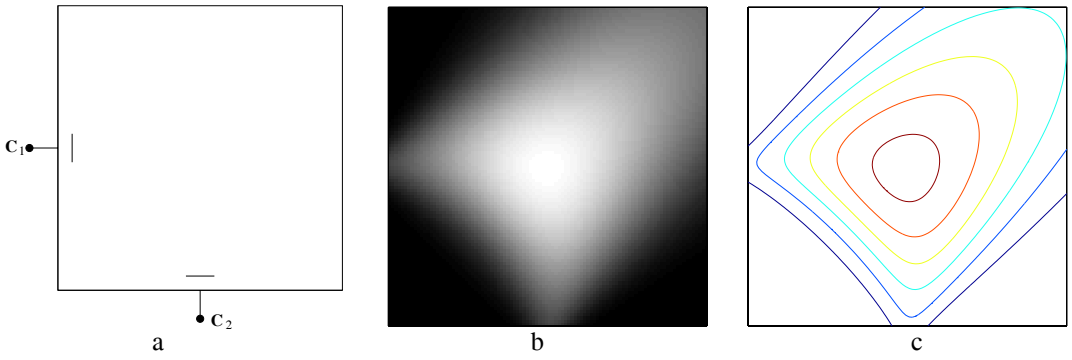

Fig. 12.7. PDF **for a triangulated point.** *(a) The camera configuration. There are two line cameras with centres at* $\mathbf{C}_1$ *and* $\mathbf{C}_2$*. The image lines are the left and lower edge of the square. The bar indicates the* 2σ *range of the noise. The plots show the* PDF *for a triangulated point computed from the two perspective projections. A large noise variance* σ^2 *is chosen to emphasize the effect. (b) The* PDF *shown as an image with white representing a higher value. (b) A contour plot of the* PDF*. Note that the* PDF *is not a Gaussian.*

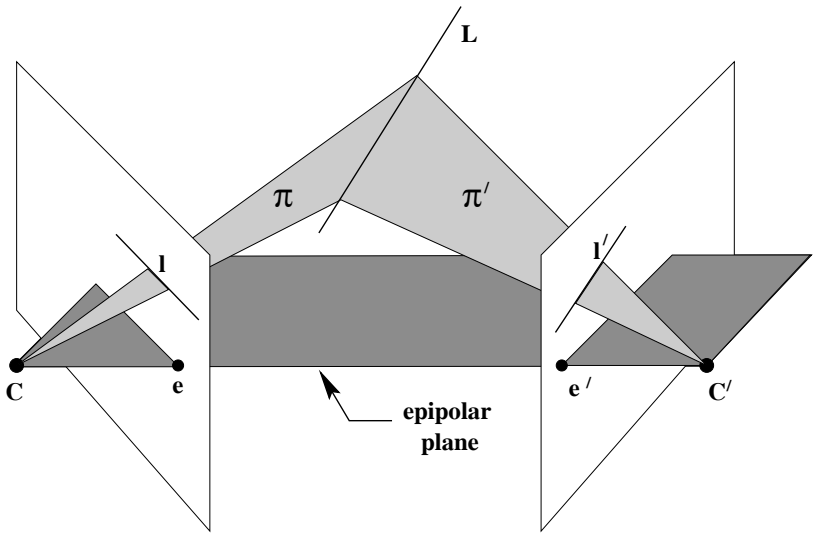

Fig. 12.8. **Line reconstruction.** *The image lines* l, l' *back-project to planes* $\boldsymbol{\pi}, \boldsymbol{\pi}'$ *respectively. The plane intersection determines the line* $\mathbf{L}$ *in 3-space. If the line in 3-space lies on an epipolar plane then its position in 3-space cannot be determined from its images. In this case the epipoles lie on the image lines.*

The planes defined by the lines are $\boldsymbol{\pi} = \mathrm{P}^{\mathsf{T}}\mathrm{l}$ and $\boldsymbol{\pi}' = \mathrm{P}'^{\mathsf{T}}\mathrm{l}'$. It is often quite convenient in practice to parametrize the line in 3-space by the two planes defined by the image lines, i.e. to represent the line as the 2×4 matrix

$$\mathrm{L} = \begin{bmatrix} \mathrm{l}^{\mathsf{T}}\mathrm{P} \\ \mathrm{l}'^{\mathsf{T}}\mathrm{P}' \end{bmatrix}$$

as described in the span representation of section 3.2.2(*p*68). Then for example a point $\mathbf{X}$ lies on the line if $\mathrm{L}\mathbf{X} = \mathbf{0}$.

In the case of corresponding points the pre-image (i.e. the point in 3-space that projects to the image points) is over-determined since there are four measurements

on the three degrees of freedom of the 3-space point. In contrast in the case of lines the pre-image is exactly determined because a line in 3-space has four degrees of freedom, and the image line provides two measurements in each view. Note, here we are considering the lines as infinite, and not using their endpoints.

Degeneracy. As illustrated in figure 12.8 lines in 3-space lying on epipolar planes cannot be determined from their images in two views. Such lines intersect the camera baseline. In practice, when there is measurement error, lines which are close to intersecting the baseline can be poorly localized in a reconstruction.

The degeneracy for lines is far more severe than for points: in the case of points there is a one-parameter family of points on the baseline which cannot be recovered. For lines there is a three-parameter family: one parameter for position on the baseline, and the other two for the star of lines through each point on the baseline.

Intersection of more than two planes

In later chapters (particularly chapter 15) we will be considering reconstruction from three or more views. To reconstruct the line that results from the intersection of several planes it is appropriate to proceed as follows. Represent each plane $\boldsymbol{\pi}_i$ by a 4-vector and form an $n \times 4$ matrix $\mathtt{A}$ for n planes with rows $\boldsymbol{\pi}_i^\mathsf{T}$. Let $\mathtt{A} = \mathtt{UDV}^\mathsf{T}$ be the singular value decomposition. The two columns of $\mathtt{V}$ corresponding to the two largest singular values span the best rank 2 approximation to $\mathtt{A}$ and may be used to define the line of intersection of the planes. If the planes are defined by back-projecting image lines, then the Maximum Likelihood estimate of the line $\mathbf{L}$ in 3-space is found by minimizing a geometric image distance between $\mathbf{L}$ projected into each image and the measured line in that image. This is discussed in section 16.4.1(*p*396).

12.8 Closure

It is not evident how to extend the polynomial method of triangulation to 3 or more views. However, the linear method extends in an obvious manner. More interestingly, the Sampson method also may be extended to 3 or more views, as is described in [Torr-97]. The disadvantage is that the computational cost (and also coding effort) increases noticeably with more views.

12.8.1 The literature

The optimal triangulation method was given by Hartley & Sturm [Hartley-97b].

12.8.2 Notes and exercises

(i) Derive a method for triangulation in the case of pure translational motion of the cameras. Hint, see figure 12.9. A closed form solution for the parameter θ is possible. This method was used in [Armstrong-94].

(ii) Adapt the polynomial triangulation method to a pair of affine cameras (or more generally, to cameras with the same principal plane). In this case, the fundamental matrix has a simple form, (14.1–*p*345), and the method reduces to a linear algorithm.

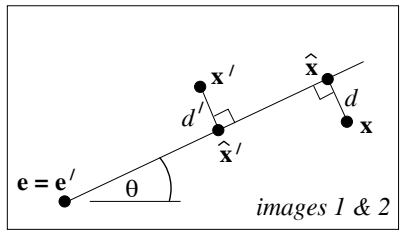

Fig. 12.9. **The epipolar geometry for pure translation.** *In this case, corresponding epipolar lines are identical (see section 11.7.1(*p293*)). The epipolar line (parametrized by θ) that minimizes $d^2 + d'^2$ can be computed directly.*

(iii) Show that the Sampson method (section 12.4) is invariant under Euclidean coordinate changes in the images (and the corresponding change in F).

(iv) Derive the analogue of the polynomial solution for triangulation in the case of a planar homography, i.e. given a measured correspondence $\mathbf{x} \leftrightarrow \mathbf{x}'$, compute points $\hat{\mathbf{x}}$ and $\hat{\mathbf{x}}'$ that minimize the function

$$\mathcal{C}(\mathbf{x}, \mathbf{x}') = d(\mathbf{x}, \hat{\mathbf{x}})^2 + d(\mathbf{x}', \hat{\mathbf{x}}')^2 \quad \text{subject to} \quad \hat{\mathbf{x}}' = \mathtt{H}\hat{\mathbf{x}}.$$

See [Sturm-97b], where it is shown that the solution is a degree 8 polynomial in one variable.