Determining the Epipolar Geometry and its Uncertainty: A Review

Zhengyou Zhang

INRIA, 2004 route des Lucioles, BP 93
F-06902 Sophia-Antipolis Cedex, France

E-mail: zzhang@sophia.inria.fr
Phone: (+33) 04 93 65 78 33
Fax: (+33) 04 93 65 78 45

Abstract

Two images of a single scene/object are related by the epipolar geometry, which can be described by a 3x3 singular matrix called the essential matrix if images' internal parameters are known, or the fundamental matrix otherwise. It captures all geometric information contained in two images, and its determination is very important in many applications such as scene modeling and vehicle navigation. This paper gives an introduction to the epipolar geometry, and provides a complete review of the current techniques for estimating the fundamental matrix and its uncertainty. A well-founded measure is proposed to compare these techniques. Projective reconstruction is also reviewed. The software which we have developed for this review is available on the Internet.

Keywords: Epipolar Geometry, Fundamental Matrix, Calibration, Reconstruction, Parameter Estimation, Robust Techniques, Uncertainty Characterization, Performance Evaluation, Software.

To appear in *The International Journal of Computer Vision*, 1997

Revised version of *INRIA Research Report N° 2927*, July 1996

Revised on August 2, 1996
based on the comments from T. Luong and A. Zisserman

List of major revisions:

1° The order used in defining the fundamental matrix has been reversed. It is now $\widetilde{\mathbf{m}}'^T\mathbf{F}\widetilde{\mathbf{m}} = 0$, instead of $\widetilde{\mathbf{m}}^T\mathbf{F}\widetilde{\mathbf{m}}' = 0$ in the previous version;

2° The methods in Sect. 3.9 (virtual parallax, linear subspace, and RANSAC) have been detailed;

3° A subsection on epipolar bands computed from the uncertainty of the estimated fundamental matrix has been added;

4° The presentation of Hartley's normalized 8-point algorithm (Sect. 3.2.5) has been reorganized. A part of the details has been moved to Appendix C.

Contents

1 Introduction

Two perspective images of a single rigid object/scene are related by the so-called *epipolar geometry*, which can be described by a 3×3 singular matrix. If the internal (intrinsic) parameters of the images (e.g., the focal length, the coordinates of the principal point, etc) are known, we can work with the *normalized image coordinates* (Faugeras 1993), and the matrix is known as the *essential matrix* (Longuet-Higgins 1981); otherwise, we have to work with the *pixel image coordinates*, and the matrix is known as the *fundamental matrix* (Luong 1992, Faugeras 1995, Luong and Faugeras 1996). It contains all geometric information that is necessary for establishing correspondences between two images, from which three-dimensional structure of the perceived scene can be inferred. In a stereovision system where the camera geometry is calibrated, it is possible to calculate such a matrix from the camera perspective projection matrices through calibration (Ayache 1991, Faugeras 1993). When the intrinsic parameters are known but the extrinsic ones (the rotation and translation between the two images) are not, the problem is known as motion and structure from motion, and has been extensively studied in Computer Vision; two excellent reviews are already available in this domain (Aggarwal and Nandhakumar 1988, Huang and Netravali 1994). We are interested here in different techniques for estimating the fundamental matrix from two uncalibrated images, i.e., the case where both the intrinsic and extrinsic parameters of the images are unknown. From this matrix, we can reconstruct a projective structure of the scene, defined up to a 4×4 matrix transformation.

The study of uncalibrated images has many important applications. The reader may wonder the usefulness of such a projective structure. We cannot obtain any metric information from a projective structure: measurements of lengths and angles do not make sense. However, a projective structure still contains rich information, such as coplanarity, collinearity, and cross ratios (ratio of ratios of distances), which is sometimes sufficient for artificial systems, such as robots, to perform tasks such as navigation and object recognition (Shashua 1994a, Zeller and Faugeras 1994, Beardsley, Zisserman and Murray 1994).

In many applications such as the reconstruction of the environment from a sequence of video images where the parameters of the video lens is submitted to continuous modification, camera calibration in the classical sense is not possible. We cannot extract any metric information, but a projective structure is still possible if the camera can be considered as a pinhole. Furthermore, if we can introduce some knowledge of the scene into the projective structure, we can obtain more specific structure of the scene. For example, by specifying a plane at infinity (in practice, we need only to specify a plane sufficiently far away), an affine structure can be computed, which preserves parallelism and ratios of distances (Quan 1993, Faugeras 1995). Hartley, Gupta and Chang (1992) first reconstruct a projective structure, and then use 8 ground reference points to obtain the Euclidean structure and the camera parameters. Mohr, Boufama and Brand (1993) embed constraints such as location of points, parallelism and vertical planes (e.g., walls) directly into a minimization procedure to determine a Euclidean structure. Robert and Faugeras (1993) show that the 3D convex hull of an object can be computed from a pair of images whose epipolar geometry is known.

If we assume that the camera parameters do not change between successive views, the projective invariants can even be used to calibrate the cameras in the classical sense without using any calibration apparatus (known as *self-calibration*) (Maybank and Faugeras 1992, Faugeras, Luong and Maybank 1992, Luong 1992, Zhang, Luong and Faugeras 1996, Enciso 1995).

Recently, we have shown (Zhang 1996a) that even in the case where images are calibrated, more reliable results can be obtained if we use the constraints arising from uncalibrated images as an intermediate step.

This paper gives an introduction to the epipolar geometry, provides a new formula of the fundamental matrix which is valid for both perspective and affine cameras, and reviews different methods reported in the literature for estimating the fundamental matrix. Furthermore, a new method is described to compare two estimations of the fundamental matrix. It is based on a measure obtained through sampling the whole visible 3D space. Projective reconstruction is also reviewed. The software called `FMatrix` which implements the reviewed methods and the software called `Fdiff` which computes the difference between two fundamental matrices are both available from my home page:

`http://www.inria.fr/robotvis/personnel/zzhang/zzhang-eng.html`

`FMatrix` detects false matches, computes the fundamental matrix and its uncertainty, and performs the projective reconstruction of the points as well. Although not reviewed, a software `AffineF` which computes the affine fundamental matrix (see Sect. 5.3) is also made available.

2 Epipolar Geometry and Problem Statement

2.1 Notation

A camera is described by the widely used pinhole model. The coordinates of a 3D point $\mathtt{M} = [x, y, z]^T$ in a world coordinate system and its retinal image coordinates $\mathbf{m} = [u, v]^T$ are related by

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{P} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} ,$$

where s is an arbitrary scale, and $\mathbf{P}$ is a 3×4 matrix, called the perspective projection matrix. Denoting the homogeneous coordinates of a vector $\mathbf{x} = [x, y, \cdots]^T$ by $\widetilde{\mathbf{x}}$, i.e., $\widetilde{\mathbf{x}} = [x, y, \cdots, 1]^T$, we have $s\widetilde{\mathbf{m}} = \mathbf{P}\widetilde{\mathtt{M}}$.

The matrix $\mathbf{P}$ can be decomposed as

$$\mathbf{P} = \mathbf{A} [\mathbf{R}\, \mathbf{t}] ,$$

where $\mathbf{A}$ is a 3×3 matrix, mapping the normalized image coordinates to the retinal image coordinates, and $(\mathbf{R}, \mathbf{t})$ is the 3D displacement (rotation and translation) from the world coordinate system to the camera coordinate system.

The quantities related to the second camera is indicated by $'$. For example, if $\mathbf{m}_i$ is a point in the first image, $\mathbf{m}'_i$ denotes its corresponding point in the second image.

A line $\mathbf{l}$ in the image passing through point $\mathbf{m} = [u, v]^T$ is described by equation $au + bv + c = 0$. Let $\mathbf{l} = [a, b, c]^T$, then the equation can be rewritten as $\mathbf{l}^T \widetilde{\mathbf{m}} = 0$ or $\widetilde{\mathbf{m}}^T \mathbf{l} = 0$. Multiplying $\mathbf{l}$ by any non-zero scalar will define the same 2D line. Thus, a 2D line is represented by a homogeneous 3D vector. The distance from point $\mathbf{m}_0 = [u_0, v_0]^T$ to line $\mathbf{l} = [a, b, c]^T$ is given by

$$d(\mathbf{m}_0, \mathbf{l}) = \frac{au_0 + bv_0 + c}{\sqrt{a^2 + b^2}} .$$

Note that we here use the *signed* distance.

Finally, we use a concise notation $\mathbf{A}^{-T} = (\mathbf{A}^{-1})^T = (\mathbf{A}^T)^{-1}$ for any invertible square matrix $\mathbf{A}$.

2.2 Epipolar Geometry and Fundamental Matrix

The epipolar geometry exists between any two camera systems. Consider the case of two cameras as shown in Fig. 1. Let C and C' be the optical centers of the first and second cameras, respectively. Given a point $\mathbf{m}$ in the first image, its corresponding point in the second image is constrained to lie on a line called the *epipolar line* of $\mathbf{m}$, denoted by $\mathbf{l}'_{\mathbf{m}}$. The line $\mathbf{l}'_{\mathbf{m}}$ is the intersection of the plane Π, defined by $\mathbf{m}$, C and C' (known as the *epipolar plane*), with the second image plane $\mathcal{I}'$. This is because image point $\mathbf{m}$ may correspond to an arbitrary point on the semi-line $C\mathtt{M}$ ($\mathtt{M}$ may be at infinity) and that the projection of $C\mathtt{M}$ on $\mathcal{I}'$ is the line $\mathbf{l}'_{\mathbf{m}}$. Furthermore, one observes that all epipolar lines of the points in the first image pass through a common point $\mathbf{e}'$, which is called the *epipole*. Epipole $\mathbf{e}'$ is the intersection of the line CC' with the image plane $\mathcal{I}'$. This can be easily understood as follows. For each point $\mathbf{m}_k$ in the first image $\mathcal{I}$, its epipolar line $\mathbf{l}'_{\mathbf{m}_k}$ in $\mathcal{I}'$ is the intersection of the plane Π^k, defined by $\mathbf{m}_k$, C and C', with image plane $\mathcal{I}'$. All epipolar planes Π^k thus form a pencil of planes containing the line CC'. They must intersect $\mathcal{I}'$ at a common point, which is $\mathbf{e}'$. Finally, one can easily see the symmetry of the epipolar geometry. The corresponding point in the first image of each point $\mathbf{m}'_k$ lying on $\mathbf{l}'_{\mathbf{m}_k}$ must lie on the epipolar line $\mathbf{l}_{\mathbf{m}'_k}$, which is the intersection of the same plane Π^k with the first image plane $\mathcal{I}$. All epipolar lines form a pencil containing the epipole $\mathbf{e}$, which is the intersection of the line CC' with the image plane $\mathcal{I}$. The symmetry leads to the following observation. If $\mathbf{m}$ (a point in $\mathcal{I}$) and $\mathbf{m}'$ (a point in $\mathcal{I}'$) correspond to a single physical point $\mathtt{M}$ in space, then $\mathbf{m}$, $\mathbf{m}'$, C and C' must lie in a single plane. This is the well-known *co-planarity constraint* in solving motion and structure from motion problems when the intrinsic parameters of the cameras are known (Longuet-Higgins 1981).

The computational significance in matching different views is that for a point in the first image, its correspondence in the second image must lie on the epipolar line in the second image, and then the search

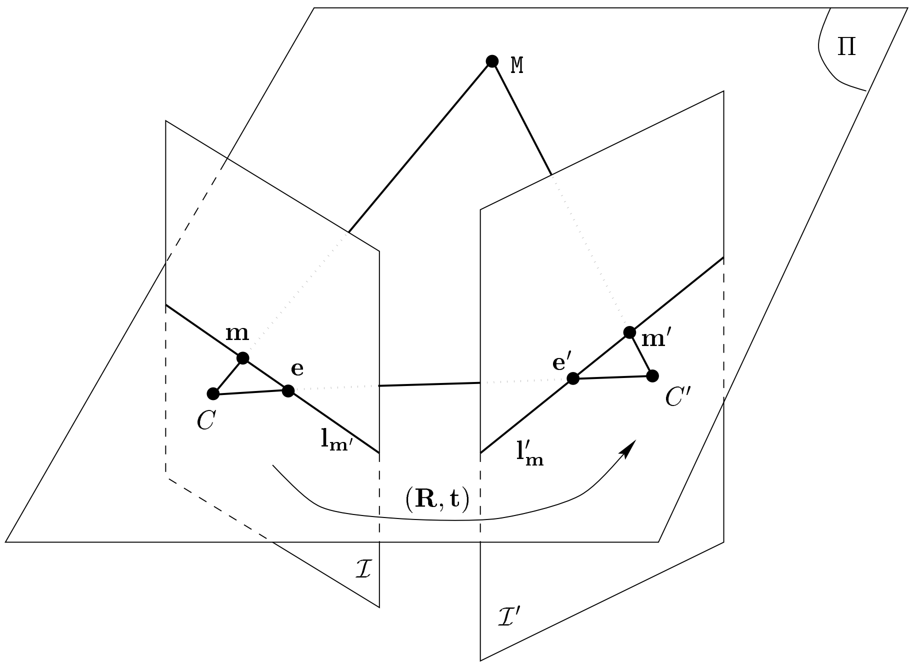

Figure 1: The epipolar geometry

space for a correspondence is reduced from 2 dimensions to 1 dimension. This is called the *epipolar constraint*. Algebraically, in order for $\mathbf{m}$ in the first image and $\mathbf{m}'$ in the second image to be matched, the following equation must be satisfied:

$$\widetilde{\mathbf{m}}'^T \mathbf{F} \widetilde{\mathbf{m}} = 0 \quad \text{with } \mathbf{F} = \mathbf{A}'^{-T} [\mathbf{t}]_\times \mathbf{R} \mathbf{A}^{-1} \;, \tag{1}$$

where $(\mathbf{R}, \mathbf{t})$ is the rigid transformation (rotation and translation) which brings points expressed in the first camera coordinate system to the second one, and $[\mathbf{t}]_\times$ is the antisymmetric matrix defined by $\mathbf{t}$ such that $[\mathbf{t}]_\times \mathbf{x} = \mathbf{t} \times \mathbf{x}$ for all 3D vector $\mathbf{x}$. This equation can be derived as follows. Without loss of generality, we assume that the world coordinate system coincides with the first camera coordinate system. From the pinhole model, we have

$$s\widetilde{\mathbf{m}} = \mathbf{A}\,[\mathbf{I}\ \mathbf{0}]\,\widetilde{\mathtt{M}} \qquad \text{and} \qquad s'\widetilde{\mathbf{m}}' = \mathbf{A}'\,[\mathbf{R}\ \mathbf{t}]\,\widetilde{\mathtt{M}} \;.$$

Eliminating $\widetilde{\mathtt{M}}$, s and s' in the above two equations, we obtain equation (1). Geometrically, $\mathbf{F}\widetilde{\mathbf{m}}$ defines the epipolar line $\mathbf{l}'_\mathbf{m}$ of point $\mathbf{m}$ in the second image. Equation (1) says no more than that the correspondence in the second image of point $\mathbf{m}$ lies on the corresponding epipolar line $\mathbf{l}'_\mathbf{m}$. Transposing (1) yields the symmetric relation from the second image to the first image: $\widetilde{\mathbf{m}}^T \mathbf{F}^T \widetilde{\mathbf{m}}' = 0$.

The 3×3 matrix $\mathbf{F}$ is called the *fundamental matrix*. Since $\det([\mathbf{t}]_\times) = 0$,

$$\det(\mathbf{F}) = 0 \;. \tag{2}$$

$\mathbf{F}$ is of rank 2. Besides, it is only defined up to a scalar factor, because if $\mathbf{F}$ is multiplied by an arbitrary scalar, equation (1) still holds. Therefore, a fundamental matrix has only seven degrees of freedom. There are only 7 independent parameters among the 9 elements of the fundamental matrix.

Convention note: We use the first camera coordinate system as the world coordinate system. In (Faugeras 1993, Xu and Zhang 1996), the second camera coordinate system is chosen as the world one. In this case, (1) becomes $\widetilde{\mathbf{m}}^T \mathbf{F}' \widetilde{\mathbf{m}}' = 0$ with $\mathbf{F}' = \mathbf{A}^{-T} [\mathbf{t}']_\times \mathbf{R}' \mathbf{A}'^{-1}$, where $(\mathbf{R}', \mathbf{t}')$ transforms points from the second camera coordinate system to the first. The relation between $(\mathbf{R}, \mathbf{t})$ and $(\mathbf{R}', \mathbf{t}')$ is given by $\mathbf{R}' = \mathbf{R}^T$, and $\mathbf{t}' = -\mathbf{R}^T \mathbf{t}$. The reader can easily verify that $\mathbf{F} = \mathbf{F}'^T$.

2.3 A General Form of Epipolar Equation for Any Projection Model

In this section we will derive a general form of epipolar equation which does not assume whether the cameras follow the perspective or affine projection model (Xu and Zhang 1996).

A point $\mathbf{m}$ in the first image is matched to a point $\mathbf{m}'$ in the second image. From the camera projection model (orthographic, weak perspective, affine, or full perspective), we have $s\widetilde{\mathbf{m}} = \mathbf{P}\widetilde{\mathtt{M}}$ and $s'\widetilde{\mathbf{m}}' = \mathbf{P}'\widetilde{\mathtt{M}}$, where $\mathbf{P}$ and $\mathbf{P}'$ are 3×4 matrices. An image point $\mathbf{m}$ defines actually an optical ray, on which every space point $\widetilde{\mathtt{M}}$ projects on the first image at $\widetilde{\mathbf{m}}$. This optical ray can be written in parametric form as

$$\widetilde{\mathtt{M}} = s\mathbf{P}^{+}\widetilde{\mathbf{m}} + \mathbf{p}^{\perp} , \tag{3}$$

where $\mathbf{P}^{+}$ is the pseudo-inverse of matrix $\mathbf{P}$:

$$\mathbf{P}^{+} = \mathbf{P}^{T}(\mathbf{P}\mathbf{P}^{T})^{-1} , \tag{4}$$

and $\mathbf{p}^{\perp}$ is any 4-vector that is perpendicular to all the *row* vectors of $\mathbf{P}$, i.e.,

$$\mathbf{P}\mathbf{p}^{\perp} = \mathbf{0} .$$

Thus, $\mathbf{p}^{\perp}$ is a null vector of $\mathbf{P}$. As a matter of fact, $\mathbf{p}^{\perp}$ indicates the position of the optical center (to which all optical rays converge). We show later how to determine $\mathbf{p}^{\perp}$. For a particular value s, equation (3) corresponds to a point on the optical ray defined by $\mathbf{m}$. Equation (3) is easily justified by projecting $\mathtt{M}$ onto the first image, which indeed gives $\mathbf{m}$.

Similarly, an image point $\mathbf{m}'$ in the second image defines also an optical ray. Requiring that the two rays to intersect in space implies that a point $\mathtt{M}$ corresponding to a particular s in (3) must project onto the second image at $\mathbf{m}'$, that is

$$s'\widetilde{\mathbf{m}}' = s\mathbf{P}'\mathbf{P}^{+}\widetilde{\mathbf{m}} + \mathbf{P}'\mathbf{p}^{\perp} .$$

Performing a cross product with $\mathbf{P}'\mathbf{p}^{\perp}$ yields

$$s'(\mathbf{P}'\mathbf{p}^{\perp}) \times \widetilde{\mathbf{m}}' = s(\mathbf{P}'\mathbf{p}^{\perp}) \times (\mathbf{P}'\mathbf{P}^{+}\widetilde{\mathbf{m}}) .$$

Eliminating s and s' by multiplying $\widetilde{\mathbf{m}}'^{T}$ from the left (equivalent to a dot product), we have

$$\widetilde{\mathbf{m}}'^{T}\mathbf{F}\widetilde{\mathbf{m}} = 0 , \tag{5}$$

where $\mathbf{F}$ is a 3×3 matrix, called *fundamental matrix*:

$$\mathbf{F} = [\mathbf{P}'\mathbf{p}^{\perp}]_{\times}\mathbf{P}'\mathbf{P}^{+} . \tag{6}$$

Since $\mathbf{p}^{\perp}$ is the optical center of the first camera, $\mathbf{P}'\mathbf{p}^{\perp}$ is actually the epipole in the second image. It can also be shown that this expression is equivalent to (1) for the full perspective projection (see Xu and Zhang 1996), but it is more general. Indeed, (1) assumes that the first 3×3 sub-matrix of $\mathbf{P}$ is invertible, and thus is only valid for full perspective projection but not for affine cameras (see Sect. 5.3), while (6) makes use of the pseudoinverse of the projection matrix, which is valid for both full perspective projection as well as affine cameras. Therefore the equation does not depend on any specific knowledge of projection model. Replacing the projection matrix in the equation by specific projection matrix for each specific projection model (e.g., orthographic, weak perspective, affine or full perspective) produces the epipolar equation for that specific projection model. See (Xu and Zhang 1996) for more details.

The vector $\mathbf{p}^{\perp}$ still needs to be determined. We first note that such a vector must exist because the difference between the row dimension and the column dimension is one, and that the row vectors are generally independent from each other. Indeed, one way to obtain $\mathbf{p}^{\perp}$ is

$$\mathbf{p}^{\perp} = (\mathbf{I} - \mathbf{P}^{+}\mathbf{P})\boldsymbol{\omega} , \tag{7}$$

where $\boldsymbol{\omega}$ is an arbitrary 4-vector. To show that $\mathbf{p}^{\perp}$ is perpendicular to each row of $\mathbf{P}$, we multiply $\mathbf{p}^{\perp}$ by $\mathbf{P}$ from the left: $\mathbf{P}\mathbf{p}^{\perp} = (\mathbf{P} - \mathbf{P}\mathbf{P}^{T}(\mathbf{P}\mathbf{P}^{T})^{-1}\mathbf{P})\boldsymbol{\omega} = \mathbf{0}$, which is indeed a zero vector. The action of $\mathbf{I} - \mathbf{P}^{+}\mathbf{P}$ is to transform an arbitrary vector to a vector that is perpendicular to every row vector of $\mathbf{P}$. If $\mathbf{P}$ is of rank 3 (which is the case for both perspective and affine cameras), then $\mathbf{p}^{\perp}$ is unique up to a scale factor.

2.4 Problem Statement

The problem considered in the sequel is the estimation of $\mathbf{F}$ from a sufficiently large set of point correspondences: $\{(\mathbf{m}_i, \mathbf{m}'_i) \mid i = 1, \ldots, n\}$, where $n \geq 7$. The point correspondences between two images can be established by a technique such as that described in (Zhang, Deriche, Faugeras and Luong 1995). We allow, however, that a fraction of the matches may be incorrectly paired, and thus the estimation techniques should be robust.

3 Techniques for Estimating the Fundamental Matrix

Let a point $\mathbf{m}_i = [u_i, v_i]^T$ in the first image be matched to a point $\mathbf{m}'_i = [u'_i, v'_i]^T$ in the second image. They must satisfy the epipolar equation (1), i.e., $\widetilde{\mathbf{m}}'^T_i \mathbf{F} \widetilde{\mathbf{m}}_i = 0$. This equation can be written as a linear and homogeneous equation in the 9 unknown coefficients of matrix $\mathbf{F}$:

$$\mathbf{u}_i^T \mathbf{f} = 0 \; , \tag{8}$$

where

$$\begin{aligned} \mathbf{u}_i &= [u_i u'_i, v_i u'_i, u'_i, u_i v'_i, v_i v'_i, v'_i, u_i, v_i, 1]^T \\ \mathbf{f} &= [F_{11}, F_{12}, F_{13}, F_{21}, F_{22}, F_{23}, F_{31}, F_{32}, F_{33}]^T \; , \end{aligned}$$

where F_{ij} is the element of $\mathbf{F}$ at row i and column j.

If we are given n point matches, by stacking (8), we have the following linear system to solve:

$$\mathbf{U}_n \mathbf{f} = \mathbf{0} \; ,$$

where

$$\mathbf{U}_n = [\mathbf{u}_1, \cdots, \mathbf{u}_n]^T \; .$$

This set of linear homogeneous equations, together with the rank constraint of the matrix $\mathbf{F}$, allow us to estimate the epipolar geometry.

3.1 Exact Solution with 7 Point Matches

As described in Sect. 2.2, a fundamental matrix $\mathbf{F}$ has only 7 degrees of freedom. Thus, 7 is the minimum number of point matches required for having a solution of the epipolar geometry.

In this case, $n = 7$ and $\operatorname{rank}(\mathbf{U}_7) = 7$. Through singular value decomposition, we obtain vectors $\mathbf{f}_1$ and $\mathbf{f}_2$ which span the null space of $\mathbf{U}_7$. The null space is a linear combination of $\mathbf{f}_1$ and $\mathbf{f}_2$, which correspond to matrices $\mathbf{F}_1$ and $\mathbf{F}_2$, respectively. Because of its homogeneity, the fundamental matrix is a one-parameter family of matrices $\alpha\mathbf{F}_1 + (1-\alpha)\mathbf{F}_2$. Since the determinant of $\mathbf{F}$ must be null, i.e.,

$$\det[\alpha\mathbf{F}_1 + (1-\alpha)\mathbf{F}_2] = 0 \; ,$$

we obtain a cubic polynomial in α. The maximum number of real solutions is 3. For each solution α, the fundamental matrix is then given by

$$\mathbf{F} = \alpha\mathbf{F}_1 + (1-\alpha)\mathbf{F}_2 \; .$$

Actually, this technique has already been used in estimating the essential matrix when 7 point matches in normalized coordinates are available (Huang and Netravali 1994). It is also used in (Hartley 1994, Torr, Beardsley and Murray 1994) for estimating the fundamental matrix.

As a matter of fact, the result that there may have three solutions given 7 matches has been known since 1800's (Hesse 1863, Sturm 1869). Sturm's algorithm (Sturm 1869) computes the epipoles and the epipolar transformation (see Sect. 2.2) from 7 point matches. It is based on the observation that the epipolar lines in the two images are related by a homography, and thus the cross-ratios of four epipolar lines is invariant. In each image, the 7 points define 7 lines going through the unknown epipole, thus providing 4 independent cross-ratios. Since these cross-ratios should remain the same in the two images, one obtains 4 cubic polynomial equations in the coordinates of the epipoles (4 independent parameters). It is shown that there may exist up to three solutions for the epipoles.

3.2 Analytic Method with 8 or More Point Matches

In practice, we are given more than 7 matches. If we ignore the rank-2 constraint, we can use a least-squares method to solve

$$\min_{\mathbf{F}} \sum_i (\widetilde{\mathbf{m}}_i'^T \mathbf{F} \widetilde{\mathbf{m}}_i)^2 \ , \tag{9}$$

which can be rewritten as:

$$\min_{\mathbf{f}} \|\mathbf{U}_n \mathbf{f}\|^2 \ . \tag{10}$$

The vector $\mathbf{f}$ is only defined up to an unknown scale factor. The trivial solution $\mathbf{f}$ to the above problem is $\mathbf{f} = \mathbf{0}$, which is not what we want. To avoid it, we need to impose some constraint on the coefficients of the fundamental matrix. Several methods are possible and are presented below. We will call them *the 8-point algorithm*, although more than 8 point matches can be used.

3.2.1 Linear Least-Squares Technique

The first method sets one of the coefficients of $\mathbf{F}$ to 1, and then solves the above problem using linear least-squares techniques. Without loss of generality, we assume that the last element of vector $\mathbf{f}$ (i.e., $f_9 = F_{33}$) is not equal to zero, and thus we can set $f_9 = -1$. This gives

$$\|\mathbf{U}_n \mathbf{f}\|^2 = \|\mathbf{U}_n' \mathbf{f}' - \mathbf{c}_9\|^2 = \mathbf{f}'^T \mathbf{U}_n'^T \mathbf{U}_n' \mathbf{f}' - 2\mathbf{c}_9^T \mathbf{U}_n' \mathbf{f}' + \mathbf{c}_9^T \mathbf{c}_9 \ ,$$

where $\mathbf{U}_n'$ is the $n \times 8$ matrix composed of the first 8 columns of $\mathbf{U}_n$, and $\mathbf{c}_9$ is the ninth column of $\mathbf{U}_n$. The solution is obtained by requiring the first derivative to be zero, i.e.,

$$\frac{\partial \|\mathbf{U}_n \mathbf{f}\|^2}{\partial \mathbf{f}'} = \mathbf{0} \ .$$

By definition of vector derivatives, $\partial(\mathbf{a}^T \mathbf{x})/\partial \mathbf{x} = \mathbf{a}$, for all vector $\mathbf{a}$. We thus have

$$2\mathbf{U}_n'^T \mathbf{U}_n' \mathbf{f}' - 2\mathbf{U}_n'^T \mathbf{c}_9 = \mathbf{0} \ , \quad \text{or} \quad \mathbf{f}' = \left(\mathbf{U}_n'^T \mathbf{U}_n'\right)^{-1} \mathbf{U}_n'^T \mathbf{c}_9 \ .$$

The problem with this method is that we do not know a priori which coefficient is not zero. If we set an element to 1 which is actually zero or much smaller than the other elements, the result will be catastrophic. A remedy is to try all nine possibilities by setting one of the nine coefficients of $\mathbf{F}$ to 1 and retain the best estimation.

3.2.2 Eigen Analysis

The second method consists in imposing a constraint on the norm of $\mathbf{f}$, and in particular we can set $\|\mathbf{f}\| = 1$. Compared to the previous method, no coefficient of $\mathbf{F}$ prevails over the others. In this case, the problem (10) becomes a classical one:

$$\min_{\mathbf{f}} \|\mathbf{U}_n \mathbf{f}\|^2 \quad \text{subject to } \|\mathbf{f}\| = 1 \ . \tag{11}$$

It can be transformed into an unconstrained minimization problem through Lagrange multipliers:

$$\min_{\mathbf{f}} \mathcal{F}(\mathbf{f}, \lambda) \ , \tag{12}$$

where

$$\mathcal{F}(\mathbf{f}, \lambda) = \|\mathbf{U}_n \mathbf{f}\|^2 + \lambda(1 - \|\mathbf{f}\|^2) \tag{13}$$

and λ is the Lagrange multiplier. By requiring the first derivative of $\mathcal{F}(\mathbf{f}, \lambda)$ with respect to $\mathbf{f}$ to be zero, we have

$$\mathbf{U}_n^T \mathbf{U}_n \mathbf{f} = \lambda \mathbf{f} \ .$$

Thus, the solution $\mathbf{f}$ must be a unit eigenvector of the 9×9 matrix $\mathbf{U}_n^T \mathbf{U}_n$ and λ is the corresponding eigenvalue. Since matrix $\mathbf{U}_n^T \mathbf{U}_n$ is symmetric and positive semi-definite, all its eigenvalues are real and positive or zero. Without loss of generality, we assume the nine eigenvalues of $\mathbf{U}_n^T \mathbf{U}_n$ are in non-increasing order:

$$\lambda_1 \geq \cdots \geq \lambda_i \geq \cdots \geq \lambda_9 \geq 0 \ .$$

We therefore have 9 potential solutions: $\lambda = \lambda_i$ for $i = 1, \ldots, 9$. Back substituting the solution to (13) gives

$$\mathcal{F}(\mathbf{f}, \lambda_i) = \lambda_i \ .$$

Since we are seeking to minimize $\mathcal{F}(\mathbf{f}, \lambda)$, the solution to (11) is evidently the unit eigenvector of matrix $\mathbf{U}_n^T \mathbf{U}_n$ associated to the *smallest* eigenvalue, i.e., λ_9.

3.2.3 Imposing the Rank-2 Constraint

The advantage of the linear criterion is that it yields an analytic solution. However, we have found that it is quite sensitive to noise, even with a large set of data points. One reason is that the rank-2 constraint (i.e., $\det \mathbf{F} = 0$) is not satisfied. We can impose this constraint a posteriori. The most convenient way is to replace the matrix $\mathbf{F}$ estimated with any of the above methods by the matrix $\hat{\mathbf{F}}$ which minimizes the Frobenius norm (see Sect. B) of $\mathbf{F} - \hat{\mathbf{F}}$ subject to the constraint $\det \hat{\mathbf{F}} = 0$. Let

$$\mathbf{F} = \mathbf{U}\mathbf{S}\mathbf{V}^T$$

be the singular value decomposition of matrix $\mathbf{F}$, where $\mathbf{S} = \operatorname{diag}(\sigma_1, \sigma_2, \sigma_3)$ is a diagonal matrix satisfying $\sigma_1 \geq \sigma_2 \geq \sigma_3$ (σ_i is the i^{th} singular value), and $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices. It can be shown that

$$\hat{\mathbf{F}} = \mathbf{U}\hat{\mathbf{S}}\mathbf{V}^T$$

with $\hat{\mathbf{S}} = \operatorname{diag}(\sigma_1, \sigma_2, 0)$ minimizes the Frobenius norm of $\mathbf{F} - \hat{\mathbf{F}}$ (see the appendix Sect. B for the proof). This method was used by Tsai and Huang (1984) in estimating the essential matrix and by Hartley (1995) in estimating the fundamental matrix.

3.2.4 Geometric Interpretation of the Linear Criterion

Another problem with the linear criterion is that the quantity we are minimizing is not physically meaningful. A physically meaningful quantity should be something measured in the image plane, because the available information (2D points) are extracted from images. One such quantity is the distance from a point $\mathbf{m}'_i$ to its corresponding epipolar line $\mathbf{l}'_i = \mathbf{F}\widetilde{\mathbf{m}}_i \equiv [l'_1, l'_2, l'_3]^T$, which is given by (see Sect. 2.1)

$$d(\mathbf{m}'_i, \mathbf{l}'_i) = \frac{\widetilde{\mathbf{m}}_i'^T \mathbf{l}'_i}{\sqrt{l_1'^2 + l_2'^2}} = \frac{1}{c'_i} \widetilde{\mathbf{m}}_i'^T \mathbf{F} \widetilde{\mathbf{m}}_i \ , \tag{14}$$

where $c'_i = \sqrt{l_1'^2 + l_2'^2}$. Thus, the criterion (9) can be rewritten as

$$\min_{\mathbf{F}} \sum_{i=1}^{n} c_i'^2 d^2(\mathbf{m}'_i, \mathbf{l}'_i) \ .$$

This means that we are minimizing not only a physical quantity $d(\mathbf{m}'_i, \mathbf{l}'_i)$, but also c'_i which is not physically meaningful. Luong (1992) shows that the linear criterion introduces a bias and tends to bring the epipoles towards the image center.

3.2.5 Normalizing Input Data

Hartley (1995) has analyzed, from a numerical computation point of view, the high instability of this linear method if pixel coordinates are directly used, and proposed to perform a simple normalization of input data prior to running the 8-point algorithm. This technique indeed produces much better results, and is summarized below.

Suppose that coordinates $\mathbf{m}_i$ in one image are replaced by $\hat{\mathbf{m}}_i = \mathbf{T}\widetilde{\mathbf{m}}_i$, and coordinates $\mathbf{m}'_i$ in the other image are replaced by $\hat{\mathbf{m}}'_i = \mathbf{T}'\widetilde{\mathbf{m}}'_i$, where $\mathbf{T}$ and $\mathbf{T}'$ are any 3×3 matrices. Substituting in the equation $\widetilde{\mathbf{m}}_i'^T \mathbf{F}\widetilde{\mathbf{m}}_i = 0$, we derive the equation $\hat{\mathbf{m}}_i'^T \mathbf{T}'^{-T}\mathbf{F}\mathbf{T}^{-1}\hat{\mathbf{m}}_i = 0$. This relation implies that $\mathbf{T}'^{-T}\mathbf{F}\mathbf{T}^{-1}$ is the fundamental matrix corresponding to the point correspondences $\hat{\mathbf{m}}_i \leftrightarrow \hat{\mathbf{m}}'_i$. Thus, an alternative method of finding the fundamental matrix is as follows:

1. Transform the image coordinates according to transformations $\hat{\mathbf{m}}_i = \mathbf{T}\widetilde{\mathbf{m}}_i$ and $\hat{\mathbf{m}}'_i = \mathbf{T}'\widetilde{\mathbf{m}}'_i$.
2. Find the fundamental matrix $\hat{\mathbf{F}}$ corresponding to the matches $\hat{\mathbf{m}}_i \leftrightarrow \hat{\mathbf{m}}'_i$.
3. Retrieve the original fundamental matrix as $\mathbf{F} = \mathbf{T}'^T\hat{\mathbf{F}}\mathbf{T}$.

The question now is how to choose the transformations $\mathbf{T}$ and $\mathbf{T}'$.

Hartley (1995) has analyzed the problem with the 8-point algorithm, which shows that its poor performance is due mainly to the poor conditioning of the problem when the pixel image coordinates are directly used (see Appendix C). Based on this, he has proposed an isotropic scaling of the input data:

1. As a first step, the points are translated so that their centroid is at the origin.
2. Then, the coordinates are scaled, so that on the average a point $\widetilde{\mathbf{m}}_i$ is of the form $\widetilde{\mathbf{m}}_i = [1, 1, 1]^T$. Such a point will lie at a distance $\sqrt{2}$ from the origin. Rather than choosing different scale factors for u and v coordinates, we choose to scale the points isotropically so that the average distance from the origin to these points is equal to $\sqrt{2}$.

Such a transformation is applied to each of the two images independently.

An alternative to the isotropic scaling is an affine transformation so that the two principal moments of the set of points are both equal to unity. However, Hartley (1995) found that the results obtained were little different from those obtained using the isotropic scaling method.

Beardsley et al. (1994) mention a normalization scheme which assumes some knowledge of camera parameters. Actually, if approximate intrinsic parameters (i.e., the intrinsic matrix $\mathbf{A}$) of a camera are available, we can apply the transformation $\mathbf{T} = \mathbf{A}^{-1}$ to obtain a "quasi-Euclidean" frame.

Boufama and Mohr (1995) use implicitly data normalization by selecting 4 points, which are largely spread in the image (i.e., most distant from each other), to form a projective basis.

3.3 Analytic Method with Rank-2 Constraint

The method described in this section is due to Faugeras (1995) which imposes the rank-2 constraint during the minimization but still yields an analytic solution. Without loss of generality, let $\mathbf{f} = [\mathbf{g}^T, f_8, f_9]^T$, where $\mathbf{g}$ is a vector containing the first seven components of $\mathbf{f}$. Let $\mathbf{c}_8$ and $\mathbf{c}_9$ be the last two column vectors of $\mathbf{U}_n$, and $\mathbf{B}$ be the $n \times 7$ matrix composed of the first seven columns of $\mathbf{U}_n$. From $\mathbf{U}_n\mathbf{f} = \mathbf{0}$, we have

$$\mathbf{B}\mathbf{g} = -f_8\mathbf{c}_8 - \mathbf{f}_9\mathbf{c}_9 \ .$$

Assume that the rank of $\mathbf{B}$ is 7, we can solve for $\mathbf{g}$ by least-squares as

$$\mathbf{g} = -f_8(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{c}_8 - f_9(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{c}_9 \ .$$

The solution depends on two free parameters f_8 and f_9. As in Sect. 3.1, we can use the constraint $\det(\mathbf{F}) = 0$, which gives a third-degree homogeneous equation in f_8 and f_9, and we can solve for their ratio. Because a third-degree equation has at least one real root, we are guaranteed to obtain at least one solution for $\mathbf{F}$. This solution is defined up to a scale factor, and we can normalize $\mathbf{f}$ such that its vector norm is equal to 1.

If there are three real roots, we choose the one that minimizes the vector norm of $\mathbf{U}_n\mathbf{f}$ subject to $\|\mathbf{f}\| = 1$. In fact, we can do the same computation for any of the 36 choices of pairs of coordinates of $\mathbf{f}$ and choose, among the possibly 108 solutions, the one that minimizes the previous vector norm.

The difference between this method and those described in Sect. 3.2 is that the latter impose the rank-2 constraint after application of the linear least-squares. We have experimented this method with a limited number of data sets, and found the results comparable with those obtained by the previous one.

3.4 Nonlinear Method Minimizing Distances of Points to Epipolar Lines

As discussed in Sect. 3.2.4, the linear method (10) does not minimize a physically meaningful quantity. A natural idea is then to minimize the distances between points and their corresponding epipolar lines: $\min_{\mathbf{F}} \sum_i d^2(\widetilde{\mathbf{m}}'_i, \mathbf{F}\widetilde{\mathbf{m}}_i)$, where $d(\cdot,\cdot)$ is given by (14). However, unlike the case of the linear criterion, the two images do not play a symmetric role. This is because the above criterion determines only the epipolar lines in the second image. As we have seen in Sect. 2.2, by exchanging the role of the two images, the fundamental matrix is changed to its transpose. To avoid the inconsistency of the epipolar geometry between the two images, we minimize the following criterion

$$\min_{\mathbf{F}} \sum_i \left(d^2(\widetilde{\mathbf{m}}'_i, \mathbf{F}\widetilde{\mathbf{m}}_i) + d^2(\widetilde{\mathbf{m}}_i, \mathbf{F}^T\widetilde{\mathbf{m}}'_i) \right) , \tag{15}$$

which operates simultaneously in the two images.

Let $\mathbf{l}'_i = \mathbf{F}\widetilde{\mathbf{m}}_i \equiv [l'_1, l'_2, l'_3]^T$ and $\mathbf{l}_i = \mathbf{F}^T\widetilde{\mathbf{m}}'_i \equiv [l_1, l_2, l_3]^T$. Using (14) and the fact that $\widetilde{\mathbf{m}}'^T_i\mathbf{F}\widetilde{\mathbf{m}}_i = \widetilde{\mathbf{m}}^T_i\mathbf{F}^T\widetilde{\mathbf{m}}'_i$, the criterion (15) can be rewritten as:

$$\min_{\mathbf{F}} \sum_i \; w_i^2 (\widetilde{\mathbf{m}}'^T_i \mathbf{F}\widetilde{\mathbf{m}}_i)^2 , \tag{16}$$

where

$$w_i = \left(\frac{1}{l_1^2 + l_2^2} + \frac{1}{l'^2_1 + l'^2_2} \right)^{1/2} = \left(\frac{l_1^2 + l_2^2 + l'^2_1 + l'^2_2}{(l_1^2 + l_2^2)(l'^2_1 + l'^2_2)} \right)^{1/2} .$$

We now present two methods for solving this problem.

3.4.1 Iterative Linear Method

The similarity between (16) and (9) leads us to solve the above problem by a *weighted* linear least-squares technique. Indeed, if we can compute the weight w_i for each point match, the corresponding linear equation can be multiplied by w_i (which is equivalent to replacing $\mathbf{u}_i$ in (8) by $w_i\mathbf{u}_i$), and exactly the same 8-point algorithm can be run to estimate the fundamental matrix, which minimizes (16).

The problem is that the weights w_i depends themselves on the fundamental matrix. To overcome this difficulty, we apply an iterative linear method. We first assume that all $w_i = 1$ and run the 8-point algorithm to obtain an initial estimation of the fundamental matrix. The weights w_i are then computed from this initial solution. The weighted linear least-squares is then run for an improved solution. This procedure can be repeated several times.

Although this algorithm is simple to implement and minimizes a physical quantity, our experience shows that there is no significant improvement compared to the original linear method. The main reason is that the rank-2 constraint of the fundamental matrix is not taken into account.

3.4.2 Nonlinear Minimization in Parameter Space

From the above discussions, it is clear that the right thing to do is to search for a matrix among the 3×3 matrices of rank 2 which minimizes (16). There are several possible parameterizations for the fundamental matrix (Luong 1992), e.g., we can express one row (or column) of the fundamental matrix as the linear combination of the other two rows (or columns). The parameterization described below is based directly on the parameters of the epipolar transformation (see Sect. 2.2).

Parameterization of fundamental matrix. Denoting the columns of $\mathbf{F}$ by the vectors $\mathbf{c}_1$, $\mathbf{c}_2$ and $\mathbf{c}_3$, we have:

$$
\begin{array}{rcl}
\text{rank}(\mathbf{F}) & = & 2 \\
& \Longleftrightarrow & \\
(\exists j_0, j_1, j_2 \in [1,3]) \ (\exists \lambda_1, \lambda_2 \in \mathcal{R}), & & \mathbf{c}_{j_0} + \lambda_1 \mathbf{c}_{j_1} + \lambda_2 \mathbf{c}_{j_2} = 0 \qquad (17) \\
(\nexists \lambda \in \mathcal{R}), & & \mathbf{c}_{j_1} + \lambda \mathbf{c}_{j_2} = 0 \ . \qquad (18)
\end{array}
$$

Condition (18), as a non-existence condition, cannot be expressed by a parameterization: we shall only keep condition (17) and so extend the parameterized set to all the 3×3-matrices of rank strictly less than 3. Indeed, the rank-2 matrices of, for example, the following forms:

$$
\begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \lambda \mathbf{c}_2 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \mathbf{c}_1 & \mathbf{0}_3 & \mathbf{c}_3 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \mathbf{0}_3 \end{bmatrix}
$$

do not have any parameterization if we take $j_0 = 1$. A parameterization of $\mathbf{F}$ is then given by $(\mathbf{c}_{j_1}, \mathbf{c}_{j_2}, \lambda_1, \lambda_2)$. This parameterization implies to divide the parameterized set among three maps, corresponding to $j_0 = 1$, $j_0 = 2$ and $j_0 = 3$.

If we construct a 3-vector such that λ_1 and λ_2 are the ${j_1}^{\text{th}}$ and ${j_2}^{\text{th}}$ coordinates and 1 is the ${j_0}^{\text{th}}$ coordinate, then it is obvious that this vector is the eigenvector of $\mathbf{F}$, and is thus the epipole in the case of the fundamental matrix. Using such a parameterization implies to compute directly the epipole which is often a useful quantity, instead of the matrix itself.

To make the problem symmetrical and since the epipole in the other image is also worth being computed, the same decomposition as for the columns is used for the rows, which now divides the parameterized set into 9 maps, corresponding to the choice of a column and a row as linear combinations of the two columns and two rows left. A parameterization of the matrix is then formed by the two coordinates x and y of the first epipole, the two coordinates x' and y' of the second epipole and the four elements a, b, c and d left by $\mathbf{c}_{i_1}$, $\mathbf{c}_{i_2}$, $\mathbf{l}_{j_1}$ and $\mathbf{l}_{j_2}$, which in turn parameterize the epipolar transformation mapping an epipolar line of the second image to its corresponding epipolar line in the first image. In that way, the matrix is written, for example, for $i_0 = 3$ and $j_0 = 3$:

$$
\mathbf{F} = \begin{bmatrix} a & b & -ax - by \\ c & d & -cx - dy \\ -ax' - cy' & -bx' - dy' & (ax + by)x' + (cx + dy)y' \end{bmatrix} \ . \tag{19}
$$

At last, to take into account the fact that the fundamental matrix is defined only up to a scale factor, the matrix is normalized by dividing the four elements (a, b, c, d) by the largest in absolute value. We have thus in total 36 maps to parameterize the fundamental matrix.

Choosing the best map. Giving a matrix $\mathbf{F}$ and the epipoles, or an approximation to it, we must be able to choose, among the different maps of the parameterization, the most suitable for $\mathbf{F}$. Denoting by $\mathbf{f}_{i_0 j_0}$ the vector of the elements of $\mathbf{F}$ once decomposed as in equation (19), i_0 and j_0 are chosen in order to maximize the rank of the 9×8 Jacobian matrix:

$$
\mathbf{J} = \frac{d\mathbf{f}_{i_0 j_0}}{d\mathbf{p}} \quad \text{where} \quad \mathbf{p} = [x, y, x', y', a, b, c, d]^T \ . \tag{20}
$$

This is done by maximizing the norm of the vector whose coordinates are the determinants of the nine 8×8 submatrices of $\mathbf{J}$. An easy calculation shows that this norm is equal to

$$
(ad - bc)^2 \sqrt{x^2 + y^2 + 1} \sqrt{x'^2 + y'^2 + 1} \ .
$$

At the expense of dealing with different maps, the above parameterization works equally well whether the epipoles are at infinity or not. This is not the case with the original proposition in (Luong 1992). More details can be found in (Csurka, Zeller, Zhang and Faugeras 1996).

Minimization. The minimization of (16) can now be performed by any minimization procedure. The Levenberg-Marquardt method (as implemented in `MINPACK` from `NETLIB` (More 1977) and in the Numeric Recipes in C (Press, Flannery, Teukolsky and Vetterling 1988)) is used in our program. During the process of minimization, the parameterization of $\mathbf{F}$ can change: The parameterization chosen for the matrix at the beginning of the process is not necessarily the most suitable for the final matrix. The nonlinear minimization method demands an initial estimate of the fundamental matrix, which is obtained by running the 8-point algorithm.

3.5 Gradient-Based Technique

Let $f_i = \widetilde{\mathbf{m}}_i'^T \mathbf{F} \widetilde{\mathbf{m}}_i$. Minimizing $\sum_i f_i^2$ does not yield a good estimation of the fundamental matrix, because the variance of each $\mathbf{f}_i$ is not the same. The least-squares technique produces an optimal solution if each term has the same variance. Therefore, we can minimize the following weighted sum of squares:

$$\min_{\mathbf{F}} \sum_i f_i^2 / \sigma_{f_i}^2 \ , \tag{21}$$

where $\sigma_{f_i}^2$ is the variance of f_i, and its computation will be given shortly. This criterion now has the desirable property: f_i/σ_{f_i} follows, under the first order approximation, the standard Gaussian distribution. In particular, all f_i/σ_{f_i} have the same variance, equal to 1. The same parameterization of the fundamental matrix as that described in the previous section is used.

Because points are extracted independently by the same algorithm, we make a reasonable assumption that the image points are corrupted by independent and identically distributed Gaussian noise, i.e., their covariance matrices are given by

$$\mathbf{\Lambda}_{\mathbf{m}_i} = \mathbf{\Lambda}_{\mathbf{m}_i'} = \sigma^2 \ \mathrm{diag}\,(1,1) \ ,$$

where σ is the noise level, which may be not known. Under the first order approximation, the variance of f_i is then given by

$$\begin{aligned} \sigma_{f_i}^2 &= \frac{\partial f_i}{\partial \mathbf{m}_i}^T \mathbf{\Lambda}_{\mathbf{m}_i} \frac{\partial f_i}{\partial \mathbf{m}_i} + \frac{\partial f_i}{\partial \mathbf{m}_i'}^T \mathbf{\Lambda}_{\mathbf{m}_i'} \frac{\partial f_i}{\partial \mathbf{m}_i'} \\ &= \sigma^2 [l_1^2 + l_2^2 + l_1'^2 + l_2'^2] \ , \end{aligned}$$

where $\mathbf{l}_i' = \mathbf{F}\widetilde{\mathbf{m}}_i \equiv [l_1', l_2', l_3']^T$ and $\mathbf{l}_i = \mathbf{F}^T \widetilde{\mathbf{m}}_i' \equiv [l_1, l_2, l_3]^T$. Since multiplying each term by a constant does not affect the minimization, the problem (21) becomes

$$\min_{\mathbf{F}} \sum_i (\widetilde{\mathbf{m}}_i'^T \mathbf{F} \widetilde{\mathbf{m}}_i)^2 / g_i^2 \ ,$$

where $g_i = \sqrt{l_1^2 + l_2^2 + l_1'^2 + l_2'^2}$ is simply the gradient of f_i. Note that g_i depends on $\mathbf{F}$.

It is shown (Luong 1992) that f_i/g_i is a first order approximation of the orthogonal distance from $(\mathbf{m}_i, \mathbf{m}_i')$ to the quadratic surface defined by $\widetilde{\mathbf{m}}'^T \mathbf{F} \widetilde{\mathbf{m}} = 0$.

3.6 Nonlinear Method Minimizing Distances Between Observation and Reprojection

If we can assume that the coordinates of the observed points are corrupted by additive noise and that the noises in different points are independent but with equal standard deviation (the same assumption as that used in the previous technique), then the maximum likelihood estimation of the fundamental matrix is obtained by minimizing the following criterion:

$$\mathcal{F}(\mathbf{f}, \mathtt{M}) = \sum_i (\|\mathbf{m}_i - \mathbf{h}(\mathbf{f}, \mathtt{M}_i)\|^2 + \|\mathbf{m}_i' - \mathbf{h}'(\mathbf{f}, \mathtt{M}_i)\|^2) \ , \tag{22}$$

where $\mathbf{f}$ represents the parameter vector of the fundamental matrix such as the one described in Sect. 3.4, $\mathtt{M} = [\mathtt{M}_1^T, \ldots, \mathtt{M}_n^T]^T$ are the structure parameters of the n points in space, while $\mathbf{h}(\mathbf{f}, \mathtt{M}_i)$ and $\mathbf{h}'(\mathbf{f}, \mathtt{M}_i)$ are the projection functions in the first and second image for a given space coordinates $\mathtt{M}_i$ and a given fundamental matrix between the two images represented by vector $\mathbf{f}$. Simply speaking, $\mathcal{F}(\mathbf{f}, \mathtt{M})$ is the sum of squared distances between observed points and the *reprojections* of the corresponding points in space. This implies that we estimate not only the fundamental matrix but also the structure parameters of the points in space. The estimation of the structure parameters, or *3D reconstruction*, in the uncalibrated case is an important subject and needs a separate section to describe it in sufficient details (see Sect. A). In the remaining subsection, we assume that there is a procedure available for 3D reconstruction.

A generalization to (22) is to take into account different uncertainties, if available, in the image points. If a point $\mathbf{m}_i$ is assumed to be corrupted by a Gaussian noise with mean zero and covariance matrix $\mathbf{\Lambda}_{\mathbf{m}_i}$ (a 2×2 symmetric positive-definite matrix), then the maximum likelihood estimation of the fundamental matrix is obtained by minimizing the following criterion:

$$\mathcal{F}(\mathbf{f}, \mathtt{M}) = \sum_i \left(\Delta\mathbf{m}_i^T \mathbf{\Lambda}_{\mathbf{m}_i}^{-1} \Delta\mathbf{m}_i + \Delta\mathbf{m}_i'^T \mathbf{\Lambda}_{\mathbf{m}_i'}^{-1} \Delta\mathbf{m}_i' \right)$$

with

$$\Delta\mathbf{m}_i = \mathbf{m}_i - \mathbf{h}(\mathbf{f}, \mathtt{M}_i) \quad \text{and} \quad \Delta\mathbf{m}_i' = \mathbf{m}_i' - \mathbf{h}'(\mathbf{f}, \mathtt{M}_i) \; .$$

Here we still assume that the noises in different points are independent, which is quite reasonable.

When the number of points n is large, the nonlinear minimization of $\mathcal{F}(\mathbf{f}, \mathtt{M})$ should be carried out in a huge parameter space ($3n+7$ dimensions because each space point has 3 degrees of freedom), and the computation is very expensive. As a matter of fact, the structure of each point can be estimated independently given an estimate of the fundamental matrix. We thus conduct the optimization of the structure parameters *in each optimization iteration* for the parameters of the fundamental matrix, that is:

$$\min_{\mathbf{f}} \left\{ \sum_i \min_{\mathtt{M}_i} \left(\|\mathbf{m}_i - \mathbf{h}(\mathbf{f}, \mathtt{M}_i)\|^2 + \|\mathbf{m}_i' - \mathbf{h}'(\mathbf{f}, \mathtt{M}_i)\|^2 \right) \right\} \; . \tag{23}$$

Therefore, a problem of minimization over $(3n + 7)$-D space (22) becomes a problem of minimization over 7-D space, in the latter each iteration contains n independent optimizations of 3 structure parameters. The computation is thus considerably reduced. As will be seen in Sect. A, the optimization of structure parameters is nonlinear. In order to speed up still more the computation, it can be approximated by an analytic method; when this optimization procedure converges, we then restart it with the nonlinear optimization method.

The idea underlying this method is already well known in motion and structure from motion (Faugeras 1993, Zhang 1995) and camera calibration (Faugeras 1993). Similar techniques have also been reported for uncalibrated images (Mohr, Veillon and Quan 1993, Hartley 1993). Because of the independence of the structure estimation (see last paragraph), the Jacobian matrix has a simple block structure in the Levenberg-Marquardt algorithm. Hartley (1993) exploits this property to simplify the computation of the pseudo-inverse of the Jacobian.

3.7 Robust Methods

Up to now, we assume that point matches are given. They can be obtained by techniques such as correlation and relaxation (Zhang, Deriche, Faugeras and Luong 1995). They all exploit some *heuristics* in one form or another, for example, intensity similarity or rigid/affine transformation in image plane, which are not applicable to most cases. Among the matches established, we may find two types of *outliers* due to

bad locations. In the estimation of the fundamental matrix, the location error of a point of interest is assumed to exhibit Gaussian behavior. This assumption is reasonable since the error in localization for most points of interest is small (within one or two pixels), but a few points are possibly incorrectly localized (more than three pixels). The latter points will severely degrade the accuracy of the estimation.

false matches. In the establishment of correspondences, only heuristics have been used. Because the only geometric constraint, i.e., the epipolar constraint in terms of the *fundamental matrix*, is not yet available, many matches are possibly false. These will completely spoil the estimation process, and the final estimate of the fundamental matrix will be useless.

The outliers will severely affect the precision of the fundamental matrix if we directly apply the methods described above, which are all least-squares techniques.

Least-squares estimators assume that the noise corrupting the data is of zero mean, which yields an *unbiased* parameter estimate. If the noise variance is known, a *minimum-variance* parameter estimate can be obtained by choosing appropriate weights on the data. Furthermore, least-squares estimators implicitly assume that the entire set of data can be interpreted by *only one parameter vector* of a given model. Numerous studies have been conducted, which clearly show that least-squares estimators are vulnerable to the violation of these assumptions. Sometimes even when the data contains only one bad datum, least-squares estimates may be completely perturbed. During the last three decades, many robust techniques have been proposed, which are not very sensitive to departure from the assumptions on which they depend.

Recently, computer vision researchers have paid much attention to the robustness of vision algorithms because the data are unavoidably error prone (Haralick 1986, Zhuang, Wang and Zhang 1992). Many the so-called *robust regression* methods have been proposed that are not so easily affected by outliers (Huber 1981, Rousseeuw and Leroy 1987). The reader is referred to (Rousseeuw and Leroy 1987, Chap. 1) for a review of different robust methods. The two most popular robust methods are the *M-estimators* and the *least-median-of-squares* (LMedS) method, which will be presented below. More details together with a description of other parameter estimation techniques commonly used in computer vision are provided in (Zhang 1996c). Recent works on the application of robust techniques to motion segmentation include (Torr and Murray 1993, Odobez and Bouthemy 1994, Ayer, Schroeter and Bigün 1994), and those on the recovery of the epipolar geometry include (Olsen 1992, Shapiro and Brady 1995, Torr 1995)

3.7.1 M-Estimators

Let r_i be the *residual* of the i^{th} datum, i.e., the difference between the i^{th} observation and its fitted value. The standard least-squares method tries to minimize $\sum_i r_i^2$, which is unstable if there are outliers present in the data. Outlying data give an effect so strong in the minimization that the parameters thus estimated are distorted. The M-estimators try to reduce the effect of outliers by replacing the squared residuals r_i^2 by another function of the residuals, yielding

$$\min \sum_i \rho(r_i) \; , \tag{24}$$

where ρ is a symmetric, positive-definite function with a unique minimum at zero, and is chosen to be less increasing than square. Instead of solving directly this problem, we can implement it as an iterated reweighted least-squares one. Now let us see how.

Let $\mathbf{p} = [p_1, \ldots, p_p]^T$ be the parameter vector to be estimated. The M-estimator of $\mathbf{p}$ based on the function $\rho(r_i)$ is the vector $\mathbf{p}$ which is the solution of the following p equations:

$$\sum_i \psi(r_i) \frac{\partial r_i}{\partial p_j} = 0 \; , \quad \text{for } j = 1, \ldots, p, \tag{25}$$

where the derivative $\psi(x) = \mathrm{d}\rho(x)/\mathrm{d}x$ is called the *influence function*. If now we define a *weight function*

$$w(x) = \frac{\psi(x)}{x} \; , \tag{26}$$

then Equation (25) becomes

$$\sum_i w(r_i) r_i \frac{\partial r_i}{\partial p_j} = 0 \; , \quad \text{for } j = 1, \ldots, p. \tag{27}$$

This is exactly the system of equations that we obtain if we solve the following iterated reweighted least-squares problem

$$\min \sum_i w(r_i^{(k-1)}) r_i^2 \, , \tag{28}$$

where the superscript $^{(k)}$ indicates the iteration number. The weight $w(r_i^{(k-1)})$ should be recomputed after each iteration in order to be used in the next iteration.

The influence function $\psi(x)$ measures the influence of a datum on the value of the parameter estimate. For example, for the least-squares with $\rho(x) = x^2/2$, the influence function is $\psi(x) = x$, that is, the influence of a datum on the estimate increases linearly with the size of its error, which confirms the non-robustness of the least-squares estimate. When an estimator is robust, it may be inferred that the influence of any single observation (datum) is insufficient to yield any significant offset (Rey 1983). There are several constraints that a robust M-estimator should meet:

- The first is of course to have a bounded influence function.
- The second is naturally the requirement of the robust estimator to be unique. This implies that the objective function of parameter vector $\mathbf{p}$ to be minimized should have a unique minimum. This requires that *the individual ρ-function is* convex *in variable* $\mathbf{p}$. This is necessary because only requiring a ρ-function to have a unique minimum is not sufficient. This is the case with maxima when considering mixture distribution; the sum of unimodal probability distributions is very often multimodal. The convexity constraint is equivalent to imposing that $\frac{\partial^2 \rho(.)}{\partial \mathbf{p}^2}$ is non-negative definite.
- The third one is a practical requirement. Whenever $\frac{\partial^2 \rho(.)}{\partial \mathbf{p}^2}$ is singular, the objective should have a gradient, i.e., $\frac{\partial \rho(.)}{\partial \mathbf{p}} \neq \mathbf{0}$. This avoids having to search through the complete parameter space.

There are a number of different M-estimators proposed in the literature. The reader is referred to (Zhang 1996c) for a comprehensive review.

It seems difficult to select a ρ-function for general use without being rather arbitrary. The result reported in Sect. 4 uses Tukey function:

$$\rho(r_i) = \begin{cases} \dfrac{c^2}{6}\left(1 - \left[1 - (\dfrac{r_i}{c\sigma})^2\right]^3\right) & \text{if } |r_i| \le c\sigma \\ (c^2/6) & \text{otherwise,} \end{cases}$$

where σ is some estimated standard deviation of errors, and $c = 4.6851$ is the tuning constant. The corresponding weight function is

$$w_i = \begin{cases} [1 - (x/c)^2]^2 & \text{if } |r_i| \le c\sigma \\ 0 & \text{otherwise.} \end{cases}$$

Another commonly used function is the following tri-weight one:

$$w_i = \begin{cases} 1 & |r_i| \le \sigma \\ \sigma/|r_i| & \sigma < |r_i| \le 3\sigma \\ 0 & 3\sigma < |r_i| \, . \end{cases}$$

In (Olsen 1992, Luong 1992), this weight function was used for the estimation of the epipolar geometry.

Inherent in the different M-estimators is the simultaneous estimation of σ, the standard deviation of the residual errors. If we can make a good estimate of the standard deviation of the errors of good data (inliers), then data whose error is larger than a certain number of standard deviations can be considered as outliers. Thus, the estimation of σ itself should be robust. The results of the M-estimators will depend on the method used to compute it. The *robust standard deviation* estimate is related to the median of the absolute values of the residuals, and is given by

$$\hat{\sigma} = 1.4826[1 + 5/(n-p)] \ \operatorname*{median}_i |r_i| \ . \tag{29}$$

The constant 1.4826 is a coefficient to achieve the same efficiency as a least-squares in the presence of only Gaussian noise (actually, the median of the absolute values of random numbers sampled from the Gaussian normal distribution $N(0,1)$ is equal to $\Phi^{-1}(\frac{3}{4}) \approx 1/1.4826$); $5/(n-p)$ (where n is the size of the data set and p is the dimension of the parameter vector) is to compensate the effect of a small set of data. The reader is referred to (Rousseeuw and Leroy 1987, page 202) for the details of these magic numbers.

Our experience shows that M-estimators are robust to outliers due to bad localization. They are, however, not robust to false matches, because they depend heavily on the initial guess, which is usually obtained by least-squares. This leads us to use other more robust techniques.

3.7.2 Least Median of Squares (LMedS)

The LMedS method estimates the parameters by solving the nonlinear minimization problem:

$$\min \ \operatorname*{median}_i \ r_i^2 \ .$$

That is, the estimator must yield the smallest value for the median of squared residuals computed for the entire data set. It turns out that this method is very robust to false matches as well as outliers due to bad localization. Unlike the M-estimators, however, the LMedS problem cannot be reduced to a weighted least-squares problem. It is probably impossible to write down a straightforward formula for the LMedS estimator. It must be solved by a search in the space of possible estimates generated from the data. Since this space is too large, only a randomly chosen subset of data can be analyzed. The algorithm which we have implemented (the original version was described in (Zhang, Deriche, Luong and Faugeras 1994, Deriche, Zhang, Luong and Faugeras 1994, Zhang, Deriche, Faugeras and Luong 1995)) for robustly estimating the fundamental matrix follows the one structured in (Rousseeuw and Leroy 1987, Chap. 5), as outlined below.

Given n point correspondences: $\{(\mathbf{m}_i, \mathbf{m}'_i) | i = 1, \ldots, n\}$, we proceed the following steps:

1. A Monte Carlo type technique is used to draw m random subsamples of $p = 7$ different point correspondences (recall that 7 is the minimum number to determine the epipolar geometry).

2. For each subsample, indexed by J, we use the technique described in Sect. 3.1 to compute the fundamental matrix $\mathbf{F}_J$. We may have at most 3 solutions.

3. For each $\mathbf{F}_J$, we can determine the median of the squared residuals, denoted by M_J, with respect to the whole set of point correspondences, i.e.,

$$M_J = \operatorname*{median}_{i=1,\ldots,n} [d^2(\widetilde{\mathbf{m}}'_i, \mathbf{F}_J \widetilde{\mathbf{m}}_i) + d^2(\widetilde{\mathbf{m}}_i, \mathbf{F}_J^T \widetilde{\mathbf{m}}'_i)] \ .$$

   Here, the distances between points and epipolar lines are used, but we can use other error measures.

4. Retain the estimate $\mathbf{F}_J$ for which M_J is minimal among all m M_J's.

The question now is: *How do we determine* m *?* A subsample is "good" if it consists of p good correspondences. Assuming that the whole set of correspondences may contain up to a fraction ε of outliers, the probability that at least one of the m subsamples is good is given by

$$P = 1 - [1 - (1-\varepsilon)^p]^m \ . \tag{30}$$

By requiring that P must be near 1, one can determine m for given values of p and ε:

$$m = \frac{\log(1-P)}{\log[1-(1-\varepsilon)^p]} \ .$$

In our implementation, we assume $\varepsilon = 40\%$ and require $P = 0.99$, thus $m = 163$. Note that the algorithm can be speeded up considerably by means of parallel computing, because the processing for each subsample can be done independently.

As noted in (Rousseeuw and Leroy 1987), the LMedS *efficiency* is poor in the presence of Gaussian noise. The efficiency of a method is defined as the ratio between the lowest achievable variance for the estimated parameters and the actual variance provided by the given method. To compensate for this deficiency, we further carry out a weighted least-squares procedure. The *robust standard deviation* estimate is given by (29), that is,

$$\hat{\sigma} = 1.4826[1 + 5/(n - p)]\sqrt{M_J} ,$$

where M_J is the minimal median estimated by the LMedS. Based on $\hat{\sigma}$, we can assign a weight for each correspondence:

$$w_i = \begin{cases} 1 & \text{if } r_i^2 \le (2.5\hat{\sigma})^2 \\ 0 & \text{otherwise ,} \end{cases}$$

where

$$r_i^2 = d^2(\widetilde{\mathbf{m}}_i', \mathbf{F}\widetilde{\mathbf{m}}_i) + d^2(\widetilde{\mathbf{m}}_i, \mathbf{F}^T\widetilde{\mathbf{m}}_i') .$$

The correspondences having $w_i = 0$ are outliers and should not be further taken into account. We thus conduct an additional step:

5. Refine the fundamental matrix $\mathbf{F}$ by solving the weighted least-squares problem:

$$\min \sum_i w_i r_i^2 .$$

The fundamental matrix is now robustly and accurately estimated because outliers have been detected and discarded by the LMedS method.

As said previously, computational efficiency of the LMedS method can be achieved by applying a Monte-Carlo type technique. However, the seven points of a subsample thus generated may be very close to each other. Such a situation should be avoided because the estimation of the epipolar geometry from such points is highly instable and the result is useless. It is a waste of time to evaluate such a subsample. In order to achieve higher stability and efficiency, we develop a *regularly random selection method* based on bucketing techniques, which works as follows. We first calculate the `min` and `max` of the coordinates of the points in the first image. The region is then evenly divided into $b \times b$ buckets (see Fig. 2). In our implementation, $b = 8$. To each bucket is attached a set of points, and indirectly a set of matches, which fall in it. The buckets having no matches attached are excluded. To generate a subsample of 7 points, we first randomly select 7 mutually different buckets, and then randomly choose one match in each selected bucket.

One question remains: How many subsamples are required? If we assume that bad matches are uniformly distributed in space, and if each bucket has the same number of matches and the random selection is uniform, the formula (30) still holds. However, the number of matches in one bucket may be quite different from that in another. As a result, a match belonging to a bucket having fewer matches has a higher probability to be selected. It is thus preferred that a bucket having many matches has a higher probability to be selected than a bucket having few matches, in order for each match to have almost the same probability to be selected. This can be realized by the following procedure. If we have in total l buckets, we divide range [0 1] into l intervals such that the width of the i^{th} interval is equal to $n_i / \sum_i n_i$, where n_i is the number of matches attached to the i^{th} bucket (see Fig. 3). During the bucket selection procedure, a number, produced by a [0 1] uniform random generator, falling in the i^{th} interval implies that the i^{th} bucket is selected.

Together with the matching technique described in (Zhang, Deriche, Faugeras and Luong 1995), we have implemented this robust method and successfully solved, in an automatic way, the matching and epipolar geometry recovery problem for different types of scenes such as indoor, rocks, road, and textured dummy scenes. The corresponding software `image-matching` has been made available on the Internet since 1994.

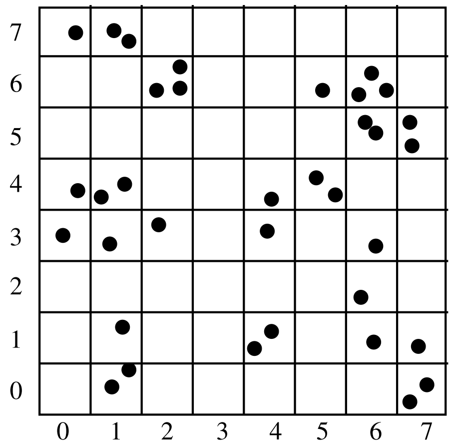

Figure 2: Illustration of a bucketing technique

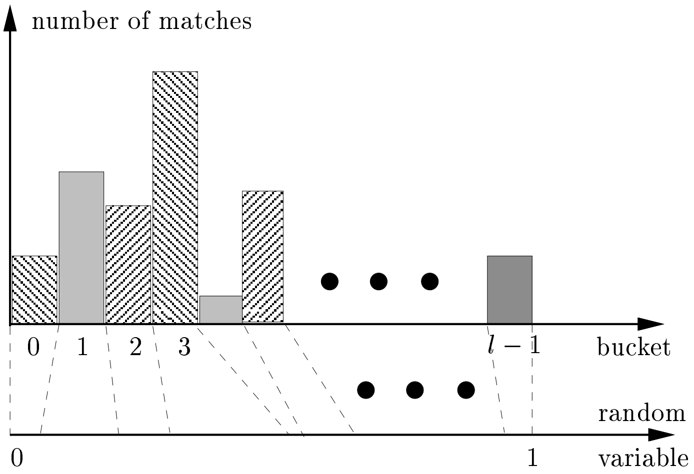

Figure 3: Interval and bucket mapping

3.8 Characterizing the Uncertainty of Fundamental Matrix

Since the data points are always corrupted by noise, and sometimes the matches are even spurious or incorrect, one should model the uncertainty of the estimated fundamental matrix in order to exploit its underlying geometric information correctly and effectively. For example, one can use the covariance of the fundamental matrix to compute the uncertainty of the projective reconstruction or the projective invariants, or to improve the results of Kruppa's equation for a better self-calibration of a camera (Zeller 1996).

In order to quantify the uncertainty related to the estimation of the fundamental matrix by the method described in the previous sections, we model the fundamental matrix as a random vector $\mathbf{f} \in \mathbb{R}^7$ (vector space of real 7-vectors) whose mean is the exact value we are looking for. Each estimation is then considered as a sample of $\mathbf{f}$ and the uncertainty is given by the covariance matrix of $\mathbf{f}$.

In the remaining of this subsection, we consider a general random vector $\mathbf{y} \in \mathbb{R}^p$, where p is the dimension of the vector space. The same discussion applies, of course, directly to the fundamental matrix. The covariance of $\mathbf{y}$ is defined by the positive symmetric matrix

$$\mathbf{\Lambda}_{\mathbf{y}} = E[(\mathbf{y} - E[\mathbf{y}])(\mathbf{y} - E[\mathbf{y}])^T] , \tag{31}$$

where $E[\mathbf{y}]$ denotes the mean of the random vector $\mathbf{y}$.

3.8.1 The Statistical Method

The statistical method consists in using the well-known large number law to approximate the mean: if we have a sufficiently large number N of samples $\mathbf{y}_i$ of a random vector $\mathbf{y}$, then $E[\mathbf{y}]$ can be approximated by the sample mean

$$E_N[\mathbf{y}_i] = \frac{1}{N}\sum_{i=1}^{N}\mathbf{y}_i \ ,$$

and $\mathbf{\Lambda_y}$ is then approximated by

$$\frac{1}{N-1}\sum_{i=1}^{N}[(\mathbf{y}_i - E_N[\mathbf{y}_i])(\mathbf{y}_i - E_N[\mathbf{y}_i])^T] \ . \tag{32}$$

A rule of thumb is that this method works reasonable well when $N > 30$. It is especially useful for simulation. For example, through simulation, we have found that the covariance of the fundamental matrix estimated by the analytical method through a first order approximation (see below) is quite good when the noise level in data points is moderate (the standard deviation is not larger than one pixel) (Csurka et al. 1996).

3.8.2 The Analytical Method

The explicit case. We now consider the case that $\mathbf{y}$ is computed from another random vector $\mathbf{x}$ of $\mathbb{R}^m$ using a C^1 function $\boldsymbol{\varphi}$:

$$\mathbf{y} = \boldsymbol{\varphi}(\mathbf{x}) \ .$$

Writing the first order Taylor expansion of $\boldsymbol{\varphi}$ in the neighborhood of $E[\mathbf{x}]$ yields

$$\boldsymbol{\varphi}(\mathbf{x}) = \boldsymbol{\varphi}(E[\mathbf{x}]) + \mathbf{D}_{\boldsymbol{\varphi}}(E[\mathbf{x}]) \cdot (\mathbf{x} - E[\mathbf{x}]) + O(\mathbf{x} - E[\mathbf{x}])^2 \ , \tag{33}$$

where $O(\mathbf{x})^2$ denotes the terms of order 2 or higher in $\mathbf{x}$, and $\mathbf{D}_{\boldsymbol{\varphi}}(\mathbf{x}) = \partial\boldsymbol{\varphi}(\mathbf{x})/\partial\mathbf{x}$ is the Jacobian matrix. Assuming that any sample of $\mathbf{x}$ is sufficiently close to $E[\mathbf{x}]$, we can approximate $\boldsymbol{\varphi}$ by the first order terms of (33) which yields:

$$\begin{aligned} E[\mathbf{y}] &\simeq \boldsymbol{\varphi}(E[\mathbf{x}]) \ , \\ \boldsymbol{\varphi}(\mathbf{x}) - \boldsymbol{\varphi}(E[\mathbf{x}]) &\simeq \mathbf{D}_{\boldsymbol{\varphi}}(E[\mathbf{x}]) \cdot (\mathbf{x} - E[\mathbf{x}]) \ . \end{aligned}$$

The first order approximation of the covariance matrix of $\mathbf{y}$ is then given in function of the covariance matrix of $\mathbf{x}$ by

$$\mathbf{\Lambda_y} = E[(\boldsymbol{\varphi}(\mathbf{x}) - \boldsymbol{\varphi}(E[\mathbf{x}]))(\boldsymbol{\varphi}(\mathbf{x}) - \boldsymbol{\varphi}(E[\mathbf{x}]))^T] = \mathbf{D}_{\boldsymbol{\varphi}}(E[\mathbf{x}])\mathbf{\Lambda_x}\mathbf{D}_{\boldsymbol{\varphi}}(E[\mathbf{x}])^T \ . \tag{34}$$

The case of an implicit function. In some cases like ours, the parameter is obtained through minimization. Therefore, $\boldsymbol{\varphi}$ is implicit and we have to make use of the well-known implicit functions theorem to obtain the following result (see Faugeras 1993, chap.6).

Proposition 1. *Let a criterion function $C : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ be a function of class C^∞, $\mathbf{x}_0 \in \mathbb{R}^m$ be the measurement vector and $\mathbf{y}_0 \in \mathbb{R}^p$ be a local minimum of $C(\mathbf{x}_0, \mathbf{z})$. If the Hessian $\mathbf{H}$ of C with respect to $\mathbf{z}$ is invertible at $(\mathbf{x}, \mathbf{z}) = (\mathbf{x}_0, \mathbf{y}_0)$ then there exists an open set U' of $\mathbb{R}^m$ containing $\mathbf{x}_0$ and an open set U'' of $\mathbb{R}^p$ containing $\mathbf{y}_0$ and a C^∞ mapping $\boldsymbol{\varphi} : \mathbb{R}^m \to \mathbb{R}^p$ such that for $(\mathbf{x}, \mathbf{y})$ in $U' \times U''$ the two relations "$\mathbf{y}$* is a local minimum of $C(\mathbf{x}, \mathbf{z})$ with respect to $\mathbf{z}$*" and $\mathbf{y} = \boldsymbol{\varphi}(\mathbf{x})$ are equivalent. Furthermore, we have the following equation:*

$$\mathbf{D}_{\boldsymbol{\varphi}}(\mathbf{x}) = -\mathbf{H}^{-1}\frac{\partial\mathbf{\Phi}}{\partial\mathbf{x}} \ , \tag{35}$$

where

$$\mathbf{\Phi} = \left(\frac{\partial C}{\partial \mathbf{z}}\right)^T \quad \textit{and} \quad \mathbf{H} = \frac{\partial \mathbf{\Phi}}{\partial \mathbf{z}} \ .$$

Taking $\mathbf{x}_0 = E[\mathbf{x}]$ and $\mathbf{y}_0 = E[\mathbf{y}]$, equation (34) then becomes

$$\mathbf{\Lambda_y} = \mathbf{H}^{-1}\frac{\partial \mathbf{\Phi}}{\partial \mathbf{x}}\mathbf{\Lambda_x}\frac{\partial \mathbf{\Phi}}{\partial \mathbf{x}}^T \mathbf{H}^{-T} \; . \tag{36}$$

The case of a sum of squares of implicit functions. Here we study the case where C is of the form:

$$\sum_{i=1}^{n} C_i^2(\mathbf{x}_i, \mathbf{z})$$

with $\mathbf{x} = [\mathbf{x}_1^T, \ldots, \mathbf{x}_i^T, \ldots, \mathbf{x}_n^T]^T$. Then, we have

$$\begin{aligned}
\mathbf{\Phi} &= 2\sum_i C_i \frac{\partial C_i}{\partial \mathbf{z}}^T \\
\mathbf{H} = \frac{\partial \mathbf{\Phi}}{\partial \mathbf{z}} &= 2\sum_i \frac{\partial C_i}{\partial \mathbf{z}}^T \frac{\partial C_i}{\partial \mathbf{z}} + 2\sum_i C_i \frac{\partial^2 C_i}{\partial \mathbf{z}^2} \; .
\end{aligned}$$

Now, it is a usual practice to neglect the terms $C_i \frac{\partial^2 C_i}{\partial \mathbf{z}^2}$ with respect to the terms $\frac{\partial C_i}{\partial \mathbf{z}}^T \frac{\partial C_i}{\partial \mathbf{z}}$ (see classical books of numerical analysis (Press et al. 1988)) and the numerical tests we did confirm that we can do this because the former is much smaller than the latter. We can then write:

$$\mathbf{H} = \frac{\partial \mathbf{\Phi}}{\partial \mathbf{z}} \approx 2\sum_i \frac{\partial C_i}{\partial \mathbf{z}}^T \frac{\partial C_i}{\partial \mathbf{z}} \; .$$

In the same way we have:

$$\frac{\partial \mathbf{\Phi}}{\partial \mathbf{x}} \approx 2\sum_i \frac{\partial C_i}{\partial \mathbf{z}}^T \frac{\partial C_i}{\partial \mathbf{x}} \; .$$

Therefore, equation (36) becomes:

$$\mathbf{\Lambda_y} = 4\mathbf{H}^{-1}\sum_{i,j} \frac{\partial C_i}{\partial \mathbf{z}}^T \frac{\partial C_i}{\partial \mathbf{x}} \mathbf{\Lambda_x} \frac{\partial C_j}{\partial \mathbf{x}}^T \frac{\partial C_j}{\partial \mathbf{z}} \mathbf{H}^{-T} \; . \tag{37}$$

Assume that the noise in $\mathbf{x}_i$ and that in $\mathbf{x}_j$ ($j \neq i$) are independent (which is quite reasonable because the points are extracted independently), then $\mathbf{\Lambda}_{\mathbf{x}_{i,j}} = E[(\mathbf{x}_i - \bar{\mathbf{x}}_i)(\mathbf{x}_j - \bar{\mathbf{x}}_j)^T] = \mathbf{0}$ and $\mathbf{\Lambda_x} = \operatorname{diag}(\mathbf{\Lambda}_{\mathbf{x}_1}, \ldots, \mathbf{\Lambda}_{\mathbf{x}_n})$. Equation (37) can then be written as

$$\mathbf{\Lambda_y} \;=\; 4\mathbf{H}^{-1}\sum_i \frac{\partial C_i}{\partial \mathbf{z}}^T \frac{\partial C_i}{\partial \mathbf{x}_i} \mathbf{\Lambda}_{\mathbf{x}_i} \frac{\partial C_i}{\partial \mathbf{x}_i}^T \frac{\partial C_i}{\partial \mathbf{z}} \mathbf{H}^{-T} \; .$$

Since $\mathbf{\Lambda}_{C_i} = \frac{\partial C_i}{\partial \mathbf{x}_i} \mathbf{\Lambda}_{\mathbf{x}_i} \frac{\partial C_i}{\partial \mathbf{x}_i}^T$ by definition (up to the first order approximation), the above equation reduces to

$$\mathbf{\Lambda_y} = 4\mathbf{H}^{-1}\sum_i \frac{\partial C_i}{\partial \mathbf{z}}^T \mathbf{\Lambda}_{C_i} \frac{\partial C_i}{\partial \mathbf{z}} \mathbf{H}^{-T}. \tag{38}$$

Considering that the mean of the value of C_i at the minimum is zero and under the somewhat strong assumption that the C_i's are independent and have identical distributed errors (Note: it is under this assumption that the solution given by the least-squares technique is optimal), we can then approximate $\mathbf{\Lambda}_{C_i}$ by its sample variance (see e.g., (Anderson 1958)):

$$\mathbf{\Lambda}_{C_i} = \frac{1}{n-p}\sum_i C_i^2 = \frac{S}{n-p} \; ,$$

where S is the value of the criterion C at the minimum, and p is the number of parameters, i.e., the dimension of $\mathbf{y}$. Although it has little influence when n is big, the inclusion of p in the formula above aims at correcting the effect of a small sample set. Indeed, for $n = p$, we can almost always find an estimate of $\mathbf{y}$ such that $C_i = 0$ for all i, and it is not meaningful to estimate the variance. Equation (38) finally becomes

$$\mathbf{\Lambda_y} = \frac{2S}{n-p}\mathbf{H}^{-1}\mathbf{H}\mathbf{H}^{-T} = \frac{2S}{n-p}\mathbf{H}^{-T} . \tag{39}$$

The case of the fundamental matrix. As explained in Sect. 3.4, $\mathbf{F}$ is computed using a sum of squares of implicit functions of n point correspondences. Thus, referring to the previous paragraph, we have $p = 7$, and the criterion function $C(\hat{\mathbf{m}}, \mathbf{f}_7)$ (where $\hat{\mathbf{m}} = [\mathbf{m}_1, \mathbf{m}'_1, \cdots, \mathbf{m}_n, \mathbf{m}'_n]^T$ and $\mathbf{f}_7$ is the vector of the seven chosen parameters for $\mathbf{F}$) is given by (15). $\mathbf{\Lambda}_{\mathbf{f}_7}$ is thus computed by (39) using the Hessian obtained as a by-product of the minimization of $C(\hat{\mathbf{m}}, \mathbf{f}_7)$.

According to (34), $\mathbf{\Lambda_F}$ is then computed from $\mathbf{\Lambda}_{\mathbf{f}_7}$:

$$\mathbf{\Lambda_F} = \frac{\partial \mathbf{F}(\mathbf{f}_7)}{\partial \mathbf{f}_7}\mathbf{\Lambda}_{\mathbf{f}_7}\frac{\partial \mathbf{F}(\mathbf{f}_7)}{\partial \mathbf{f}_7}^T . \tag{40}$$

Here, we actually consider the fundamental matrix $\mathbf{F}(\mathbf{f}_7)$ as a 9-vector composed of the 9 coefficients which are functions of the 7 parameters $\mathbf{f}_7$.

The reader is referred to (Zhang and Faugeras 1992, chap.2) for a more detailed exposition on uncertainty manipulation.

3.9 Other Techniques

To close the review section, we present two analytical techniques and one robust technique based on RANSAC.

3.9.1 Virtual Parallax Method

If two sets of image points are the projections of a plane in space (see Sect. 5.2), then they are related by a homography $\mathbf{H}$. For points not on the plane, they do not verify the homography, i.e., $\widetilde{\mathbf{m}}' \neq \rho\mathbf{H}\widetilde{\mathbf{m}}$, where ρ is an arbitrary non-zero scalar. The difference (i.e., parallax) allows us to estimate directly an epipole if the knowledge of $\mathbf{H}$ is available. Indeed, Luong and Faugeras (1996) show that the fundamental matrix and the homography is related by $\mathbf{F} = [\widetilde{\mathbf{e}}']_\times\mathbf{H}$. For a point which does not belong to the plane, $\mathbf{l}' = \widetilde{\mathbf{m}}' \times \mathbf{H}\widetilde{\mathbf{m}}$ defines an epipolar line, which provides one constraint on the epipole: $\widetilde{\mathbf{e}}'^T\mathbf{l}' = 0$. Therefore, two such points are sufficient to estimate the epipole $\mathbf{e}'$. The generate-and-test methods, see e.g., (Faugeras and Lustman 1988), can be used to detect the coplanar points.

The virtual parallax method proposed by Boufama and Mohr (1995) does not require the prior identification of a plane. To simplify the computations, without loss of generality, we can perform a change of projective coordinates in each image such that

$$\widetilde{\mathbf{m}}_1 = [1,0,0]^T,\ \widetilde{\mathbf{m}}_2 = [0,1,0]^T,\ \widetilde{\mathbf{m}}_3 = [0,0,1]^T,\ \widetilde{\mathbf{m}}_4 = [1,1,1]^T ; \tag{41}$$
$$\widetilde{\mathbf{m}}'_1 = [1,0,0]^T,\ \widetilde{\mathbf{m}}'_2 = [0,1,0]^T,\ \widetilde{\mathbf{m}}'_3 = [0,0,1]^T,\ \widetilde{\mathbf{m}}'_4 = [1,1,1]^T . \tag{42}$$

These points are chosen such that no three of them are collinear. The three first points define a plane in space. Under such choice of coordinate systems, the homography matrix such that $\widetilde{\mathbf{m}}'_i = \rho\mathbf{H}\widetilde{\mathbf{m}}_i$ $(i = 1,2,3)$ is diagonal, i.e., $\mathbf{H} = \operatorname{diag}(a,b,c)$, and depends only on two parameters. Let the epipole be $\widetilde{\mathbf{e}}' = [e'_u, e'_v, e'_t]^T$. As we have seen in the last paragraph, for each additional point $(\mathbf{m}_i, \mathbf{m}'_i)$ $(i = 4, \ldots, n)$, we have $\widetilde{\mathbf{e}}'^T(\widetilde{\mathbf{m}}'_i \times \mathbf{H}\widetilde{\mathbf{m}}_i) = 0$, i.e.,

$$v'_i e'_u c - v_i e'_u b + u_i e'_v a - u'_i e'_v c + u'_i v_i e'_t b - v'_i u_i e'_t a = 0 . \tag{43}$$

This is the basic epipolar equation based on virtual parallax. Since (a,b,c) and (e'_u, e'_v, e'_t) are defined each up to a scale factor, the above equation is polynomial of degree two in four unknowns. To simplify the problem, we make the following reparameterization. Let

$$x_1 = e'_u c,\ x_2 = e'_u b,\ x_3 = e'_v a,\ x_4 = e'_v c,\ x_5 = e'_t b, \text{ and } x_6 = e'_t a \ ,$$

which are defined up to a common scale factor. Equation (43) now becomes

$$v'_i x_1 - v_i x_2 + u_i x_3 - u'_i x_4 + u'_i v_i x_5 - v'_i u_i x_6 = 0 \ . \tag{44}$$

Unlike (43), we here have five independent variables, one more than necessary. The unknowns x_i ($i = 1, \ldots, 6$) can be solved linearly if we have five or more point matches. Thus, we need in total 8 point correspondences, like the eight-point algorithm. The original unknowns can be computed, for example, as

$$e'_u = e'_t x_2 / x_5,\ e'_v = e'_t x_3 / x_6,\ a = c x_3 / x_4, \text{ and } b = c x_2 / x_1 \ . \tag{45}$$

The fundamental matrix is finally obtained as $[\widetilde{\mathbf{e}}']_\times \operatorname{diag}(a, b, c)$, and the rank constraint is automatically satisfied. However, note that

- the computation (45) is not optimal, because each intermediate variable x_i is not used equally;
- the rank-2 constraint in the linear equation (44) is not necessarily satisfied because of the introduction of an intermediate parameter.

Therefore, the rank-2 constraint is also imposed a posteriori, similar to the 8-point algorithm (see Sect. 3.2).

The results obtained with this method depends on the choice of the four basis points. The authors indicate that a good choice is to take them largely spread in the image.

Experiments show that this method produces good results. Factors which contribute to this are the fact the dimensionality of the problem has been reduced, and the fact that the change of projective coordinates achieve a data renormalization comparable to the one described in Sect. 3.2.5.

3.9.2 Linear Subspace Method

Ponce and Genc (1996), through a change of projective coordinates, set up a set of linear constraints on one epipole using the linear subspace method proposed by Heeger and Jepson (1992). A change of projective coordinates in each image as described in (41) and (42) is performed. Furthermore, we choose the corresponding four scene points $\mathtt{M}_i (i = 1, \ldots, 4)$ and the optical center of each camera as a projective basis in space. We assign to the basis points for the first camera the following coordinates:

$$\widetilde{\mathtt{M}}_1 = [1,0,0,0]^T,\ \widetilde{\mathtt{M}}_2 = [0,1,0,0]^T,\ \widetilde{C} = [0,0,1,0]^T,\ \widetilde{\mathtt{M}}_3 = [0,0,0,1]^T,\ \widetilde{\mathtt{M}}_4 = [1,1,1,1]^T \ . \tag{46}$$

The same coordinates are assigned to the basis points for the second camera. Therefore, the camera projection matrix for the first camera is given by

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \ . \tag{47}$$

Let the coordinates of the optical center C of the first camera be $[\alpha, \beta, \gamma, 1]^T$ in the projective basis of the second camera, and let the coordinates of the four scene points remain the same in both projective bases, i.e., $\mathtt{M}'_i = \mathtt{M}_i$ ($i = 1, \ldots, 4$). Then, the coordinate transformation $\mathbf{H}$ from the projective basis of the first camera to that of the second camera is given by

$$\mathbf{H} = \begin{bmatrix} \gamma - \alpha & 0 & \alpha & 0 \\ 0 & \gamma - \beta & \beta & 0 \\ 0 & 0 & \gamma & 0 \\ 0 & 0 & 1 & \gamma - 1 \end{bmatrix} \ . \tag{48}$$

It is then a straightforward manner to obtain the projection matrix of the first camera with respect to the projective basis of the second camera:

$$\mathbf{P}' = \mathbf{P}\mathbf{H} = \begin{bmatrix} \gamma - \alpha & 0 & \alpha & 0 \\ 0 & \gamma - \beta & \beta & 0 \\ 0 & 0 & 1 & \gamma - 1 \end{bmatrix} . \tag{49}$$

According to (6), the epipolar equation is $\widetilde{\mathbf{m}}_i'^T \mathbf{F} \widetilde{\mathbf{m}}_i = 0$, while the fundamental matrix is given by $\mathbf{F} = [\mathbf{P}'\mathbf{p}^\perp]_\times \mathbf{P}'\mathbf{P}^+$. Since

$$\mathbf{p}^\perp = C = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{P}^+ = \mathbf{P}^T(\mathbf{P}\mathbf{P}^T)^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} ,$$

we obtain the fundamental matrix:

$$\mathbf{F} = [\widetilde{\mathbf{e}}']_\times \ \mathrm{diag}\,(\gamma - \alpha, \gamma - \beta, \gamma - 1) , \tag{50}$$

where $\widetilde{\mathbf{e}}' \equiv \mathbf{P}'\mathbf{p}^\perp = [\alpha, \beta, 1]^T$ is just the projection of the first optical center in the second camera, i.e., the second epipole.

Consider now the remaining point matches $\{(\mathbf{m}_i, \mathbf{m}_i') | i = 5, \ldots, n\}$, where $\widetilde{\mathbf{m}}_i = [u_i, v_i, 1]^T$ and $\widetilde{\mathbf{m}}_i' = [u_i', v_i', 1]^T$. From (50), after some simple algebraic manipulation, the epipolar equation can be rewritten as

$$\gamma \mathbf{g}_i^T \widetilde{\mathbf{e}}' = \mathbf{q}_i^T \mathbf{f} ,$$

where $\mathbf{f} = [\alpha, \beta, \alpha\beta]^T$, $\mathbf{g}_i = \widetilde{\mathbf{m}}_i' \times \widetilde{\mathbf{m}}_i = [v_i' - v_i, u_i - u_i', -v_i' u_i + u_i' v_i]^T$ and $\mathbf{q}_i = [v_i'(1 - u_i), -u_i'(1 - v_i), u_i - v_i]^T$. Consider a linear combination of the above equations. Let us define the coefficient vector $\boldsymbol{\xi} = [\xi_5, \ldots, \xi_n]^T$ and the vectors $\boldsymbol{\tau}(\boldsymbol{\xi}) = \sum_{i=5}^n \xi_i \mathbf{g}_i$ and $\boldsymbol{\chi}(\boldsymbol{\xi}) = \sum_{i=5}^n \xi_i \mathbf{q}_i$. It follows that

$$\gamma \boldsymbol{\tau}(\boldsymbol{\xi})^T \widetilde{\mathbf{e}}' = \boldsymbol{\chi}(\boldsymbol{\xi})^T \mathbf{f} . \tag{51}$$

The idea of the linear subspace is that for any value $\boldsymbol{\xi}_{\boldsymbol{\tau}}$ such that $\boldsymbol{\tau}(\boldsymbol{\xi}_{\boldsymbol{\tau}}) = \mathbf{0}$, equation (51) provides a linear constraint on $\mathbf{f}$, i.e., $\boldsymbol{\chi}(\boldsymbol{\xi}_{\boldsymbol{\tau}})^T \mathbf{f} = \mathbf{0}$, while for any value $\boldsymbol{\xi}_{\boldsymbol{\chi}}$ such that $\boldsymbol{\chi}(\boldsymbol{\xi}_{\boldsymbol{\chi}}) = \mathbf{0}$, the same equation provides a linear constraint on $\widetilde{\mathbf{e}}'$, i.e., $\boldsymbol{\tau}(\boldsymbol{\xi}_{\boldsymbol{\chi}})^T \widetilde{\mathbf{e}}' = \mathbf{0}$. Because of the particular structure of $\mathbf{g}_i$ and $\mathbf{q}_i$, it is easy to show (Ponce and Genc 1996) that the vectors $\boldsymbol{\tau}(\boldsymbol{\xi}_{\boldsymbol{\chi}})$ and $\boldsymbol{\chi}(\boldsymbol{\xi}_{\boldsymbol{\tau}})$ are both orthogonal to the vector $[1, 1, 1]^T$. Since the vectors $\boldsymbol{\tau}(\boldsymbol{\xi}_{\boldsymbol{\chi}})$ are also orthogonal to $\widetilde{\mathbf{e}}'$, they only span a one-dimensional line, and their representative vector is denoted by $\boldsymbol{\tau}_0 = [a_\tau, b_\tau, c_\tau]^T$. Likewise, the vectors $\boldsymbol{\chi}(\boldsymbol{\xi}_{\boldsymbol{\tau}})$ span a line orthogonal to both $\mathbf{f}$ and $[1, 1, 1]^T$, and their representative vector is denoted by $\boldsymbol{\chi}_0 = [a_\chi, b_\chi, c_\chi]^T$. Assume for the moment that we know $\boldsymbol{\tau}_0$ and $\boldsymbol{\chi}_0$ (their computation will be described shortly), from $[a_\tau, b_\tau, -a_\tau - b_\tau]^T \widetilde{\mathbf{e}}' = 0$ and $[a_\chi, b_\chi, -a_\chi - b_\chi]^T \mathbf{f} = 0$, the solution to the epipole is given by

$$\alpha = \frac{b_\chi}{a_\tau} \frac{a_\tau + b_\tau}{a_\chi + b_\chi} , \quad \beta = \frac{a_\chi}{b_\tau} \frac{a_\tau + b_\tau}{a_\chi + b_\chi} . \tag{52}$$

Once the epipole has been computed, the remaining parameters of the fundamental matrix can be easily computed.

We now turn to the estimation of $\boldsymbol{\tau}_0$ and $\boldsymbol{\chi}_0$. From the above discussion, we see that the set of linear combinations $\sum_{i=5}^n \xi \mathbf{g}_i$ such that $\sum_{i=5}^n \xi \mathbf{q}_i = \mathbf{0}$ is one-dimensional. Construct two $3 \times (n-4)$ matrices:

$$\mathbf{G} = [\mathbf{g}_5, \ldots, \mathbf{g}_n] \quad \text{and} \quad \mathbf{Q} = [\mathbf{q}_5, \ldots, \mathbf{q}_n] .$$

The set of vectors $\boldsymbol{\xi}$ such that $\sum_{i=5}^n \xi \mathbf{q}_i = \mathbf{0}$ is simply the null space of $\mathbf{Q}$. Let $\mathbf{Q} = \mathbf{U}_1 \mathbf{S}_1 \mathbf{V}_1^T$ be the singular value decomposition (SVD) of $\mathbf{Q}$, then the null space is formed by the rightmost $n - 4 - 3 = n - 7$ columns of $\mathbf{V}_1$, which will be denoted by $\mathbf{V}_0$. Then, the set of vectors $\sum_{i=5}^n \xi \mathbf{g}_i$ such that $\sum_{i=5}^n \xi \mathbf{q}_i = \mathbf{0}$ is thus the subspace spanned by the matrix $\mathbf{G}\mathbf{V}_0$, which is $3 \times (n-7)$. Let $\mathbf{G}\mathbf{V}_0 = \mathbf{U}_2 \mathbf{S}_2 \mathbf{V}_2^T$ be the SVD. According to our assumptions, this matrix has rank 1, thus $\boldsymbol{\tau}_0$ is the *range* of $\mathbf{GV}_0$, which is simply the leftmost column of $\mathbf{U}_2$ up to a scale factor. Vector $\boldsymbol{\chi}_0$ can be computed following the same construction by reversing the rôles of $\boldsymbol{\tau}$ and $\boldsymbol{\chi}$.

The results obtained with this method depends on the choice of the four basis points. The authors show experimentally that a good result can be obtained by trying 30 random basis choices and picking up the solution resulting the smallest epipolar distance error.

Note that although unlike the virtual parallax method, the linear subspace technique provides a linear algorithm without introducing an extraneous parameter, it is achieved in (52) by simply dropping the estimated information in c_τ and c_χ. In the presence of noise, $\boldsymbol{\tau}_0$ and $\boldsymbol{\chi}_0$ computed through singular value decomposition do not necessarily satisfy $\boldsymbol{\tau}_0^T\mathbf{1} = 0$ and $\boldsymbol{\chi}_0^T\mathbf{1} = 0$, where $\mathbf{1} = [1, 1, 1]^T$

Experiments show that this method produces good results. The same reasons as for the virtual parallax method can be used here.

3.9.3 RANSAC

RANSAC (random sample consensus) (Fischler and Bolles 1981) is a paradigm originated in the Computer Vision community for robust parameter estimation. The idea is to find, through random sampling of a minimal subset of data, the parameter set which is consistent with a subset of data as large as possible. The consistent check requires the user to supply a threshold on the errors, which reflects the a priori knowledge of the precision of the expected estimation. This technique is used by Torr (1995) to estimate the fundamental matrix. As is clear, RANSAC is very similar to LMedS both in ideas and in implementation, except that

- RANSAC needs a threshold to be set by the user for consistence checking, while the threshold is automatically computed in LMedS;
- In step 3 of the LMedS implementation described in Sect. 3.7.2, the size of the point matches which are consistent with $\mathbf{F}_J$ is computed, instead of the median of the squared residuals.

However, LMedS cannot deal with the case where the percentage of outliers is higher than 50%, while RANSAC can. Torr and Murray (1993) compared both LMedS and RANSAC. RANSAC is usually cheaper because it can exit the random sampling loop once a consistent solution is found.

If one knows that the number of outliers is more than 50%, then they can easily adapt the LMedS by using an appropriate value, say 40%, instead of using the median. (When we do this, however, the solution obtained may be not *globally* optimal if the number of outliers is less than 50%.) If there is a large set of images of the same type of scenes to be processed, one can first apply LMedS to one pair of the images in order to find an appropriate threshold, and then apply RANSAC to the remaining images because it is cheaper.

4 An Example of Fundamental Matrix Estimation With Comparison

The pair of images is a pair of calibrated stereo images (see Fig. 4). By "calibrated" is meant that the intrinsic parameters of both cameras and the displacement between them were computed off-line through stereo calibration. There are 241 point matches, which are established automatically by the technique described in (Zhang, Deriche, Faugeras and Luong 1995). Outliers have been discarded. The calibrated parameters of the cameras are of course not used, but the fundamental matrix computed from these parameters serves as a ground truth. This is shown in Fig. 5, where the four epipolar lines are displayed, corresponding, from the left to the right, to the point matches 1, 220, 0 and 183, respectively. The intersection of these lines is the epipole, which is clearly very far from the image. This is because the two cameras are placed almost in the same plane.

The epipolar geometry estimated with the linear method is shown in Fig. 6 for the same set of point matches. One can find that the epipole is now in the image, which is completely different from what we have seen with the calibrated result. If we perform a data normalization before applying the linear method, the result is considerably improved, as shown in Fig. 7. This is very close to the calibrated one.

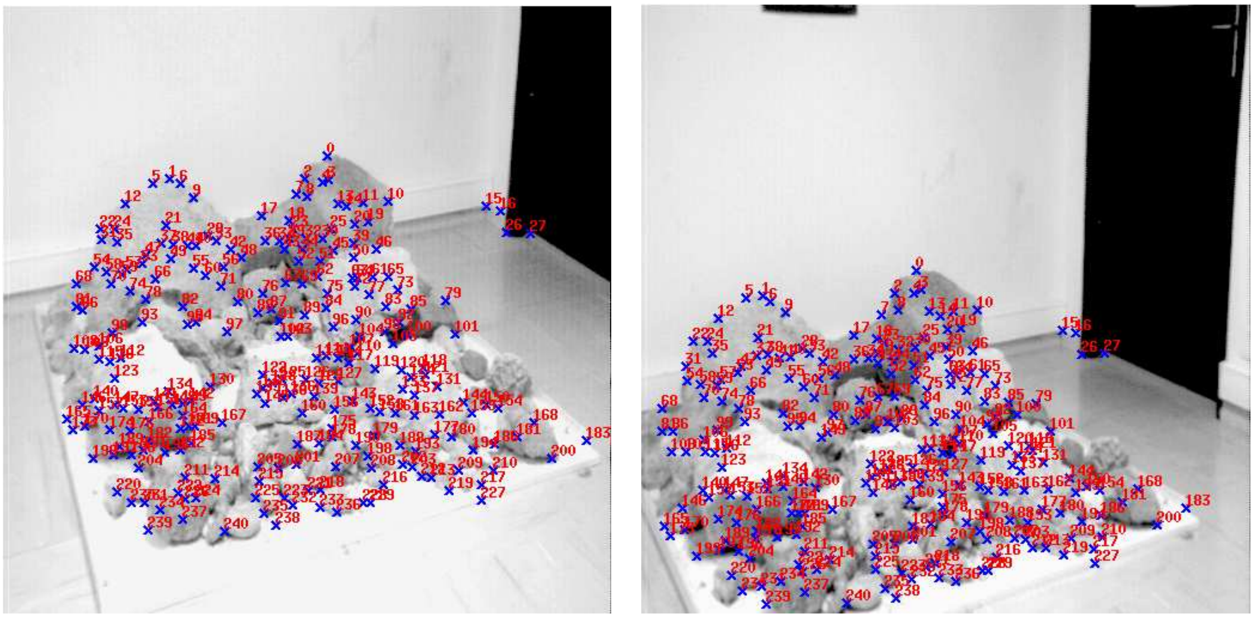

Figure 4: Image pair used for comparing different estimation techniques of the fundamental matrix

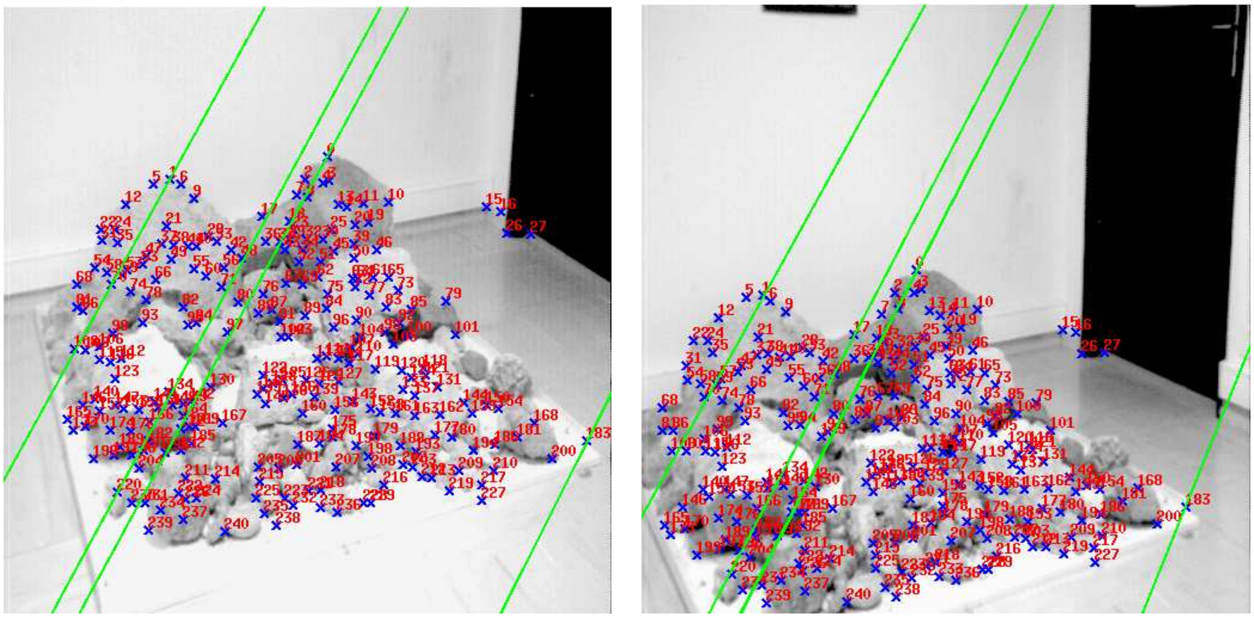

Figure 5: Epipolar geometry estimated through classical stereo calibration, which serves as the ground truth

The nonlinear method gives even better result, as shown in Fig. 8. A comparison with the "true" epipolar geometry is shown in Fig. 9. There is only a small difference in the orientation of the epipolar lines. We have also tried the normalization method followed by the nonlinear method, and the same result was obtained. Other methods have also been tested, and visually almost no difference is observed.

Quantitative results are provided in Table 1, where the elements in the first column indicates the methods used in estimating the fundamental matrix: they are respectively the classical stereo calibration (**Calib.**), the linear method with eigen analysis (**linear**), the linear method with prior data normalization (**normal.**), the nonlinear method based on minimization of distances between points and epipolar lines (**nonlinear**), the

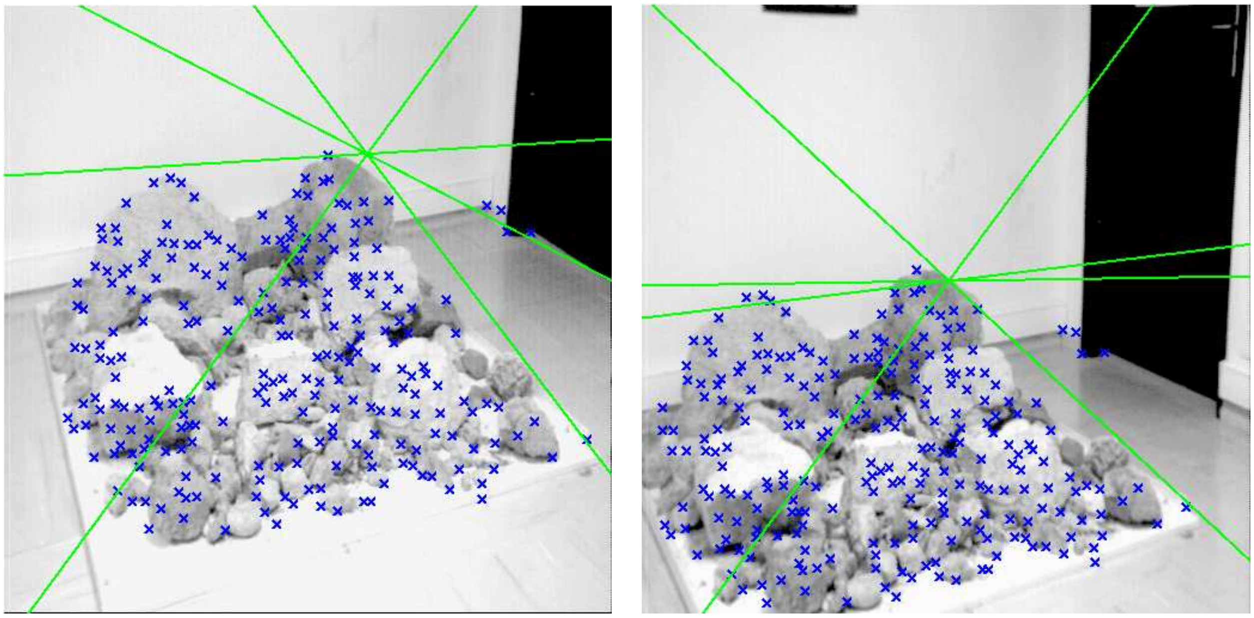

Figure 6: Epipolar geometry estimated with the linear method

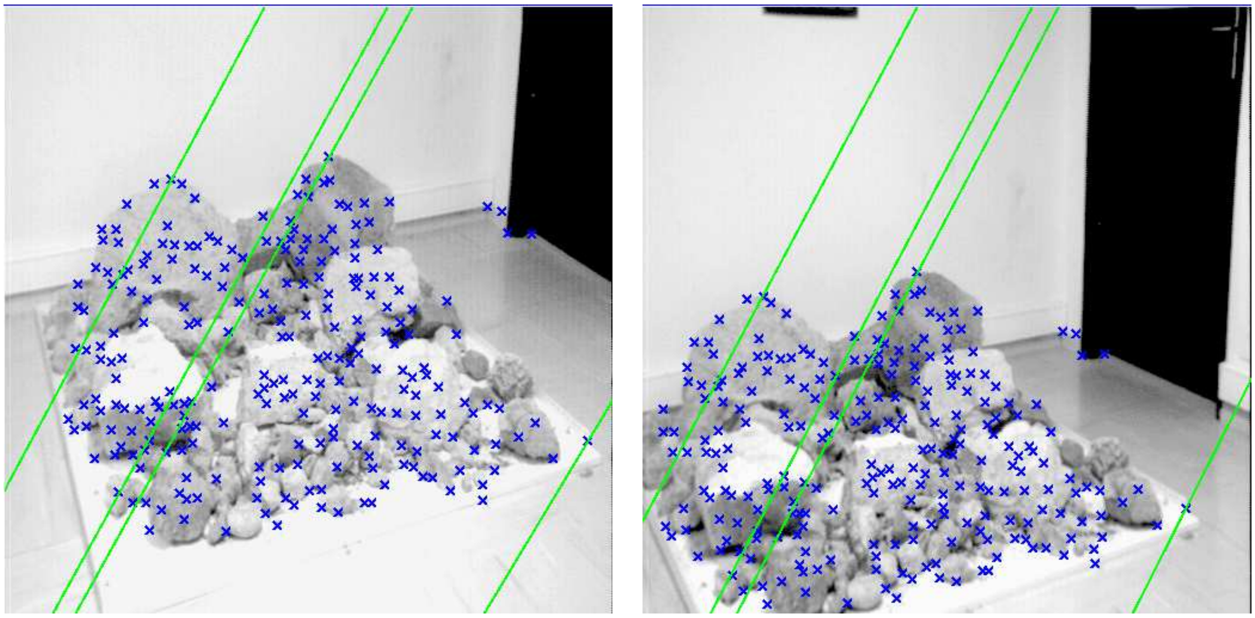

Figure 7: Epipolar geometry estimated with the linear method with prior data normalization

nonlinear method based on minimization of gradient-weighted epipolar errors (**gradient**), the M-estimator with Tukey function (**M-estim.**), the nonlinear method based on minimization of distances between observed points and reprojected ones (**reproj.**), and the LMedS technique (**LMedS**). The fundamental matrix of **Calib** is used as a reference. The second column shows the difference between the fundamental matrix estimated by each method with that of **Calib**. The difference is measured as the Frobenius norm: $\Delta\mathbf{F} = \|\mathbf{F} - \mathbf{F}_{\mathbf{Calib}}\| \times 100\%$. Since each $\mathbf{F}$ is normalized by its Frobenius norm, $\Delta\mathbf{F}$ is directly related to the angle between two unit vectors. It can be seen that although we have observed that Method **normal** has considerably improved the result of the linear method, its $\Delta\mathbf{F}$ is the largest. It seems that $\Delta\mathbf{F}$ is not

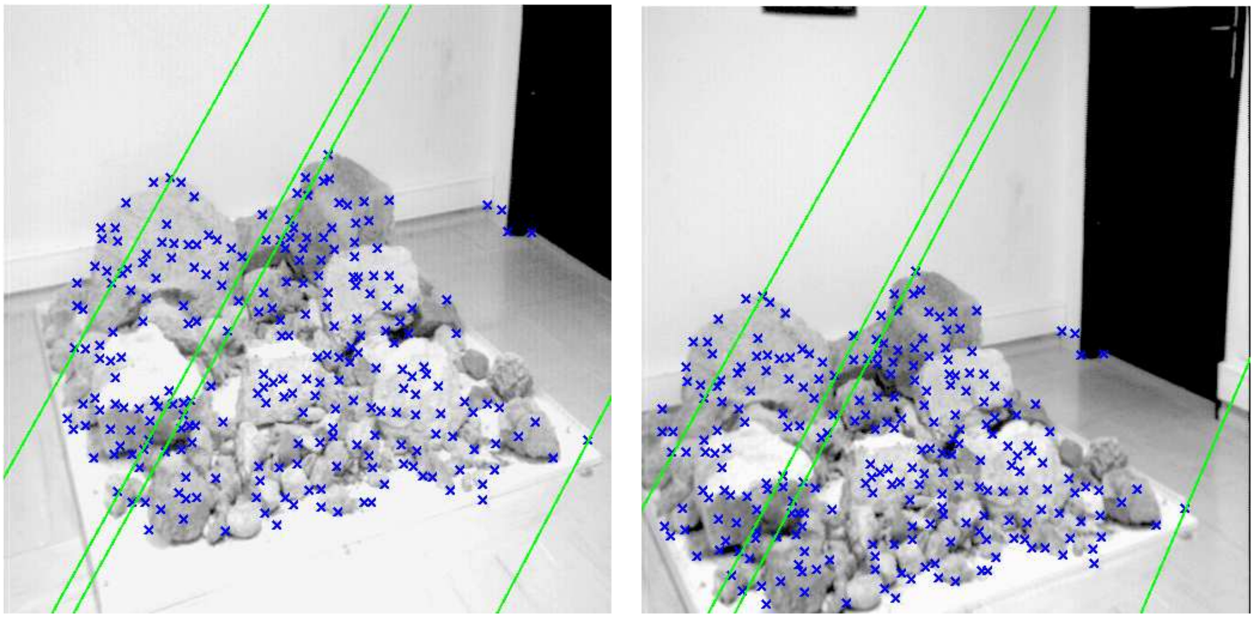

Figure 8: Epipolar geometry estimated with the nonlinear method

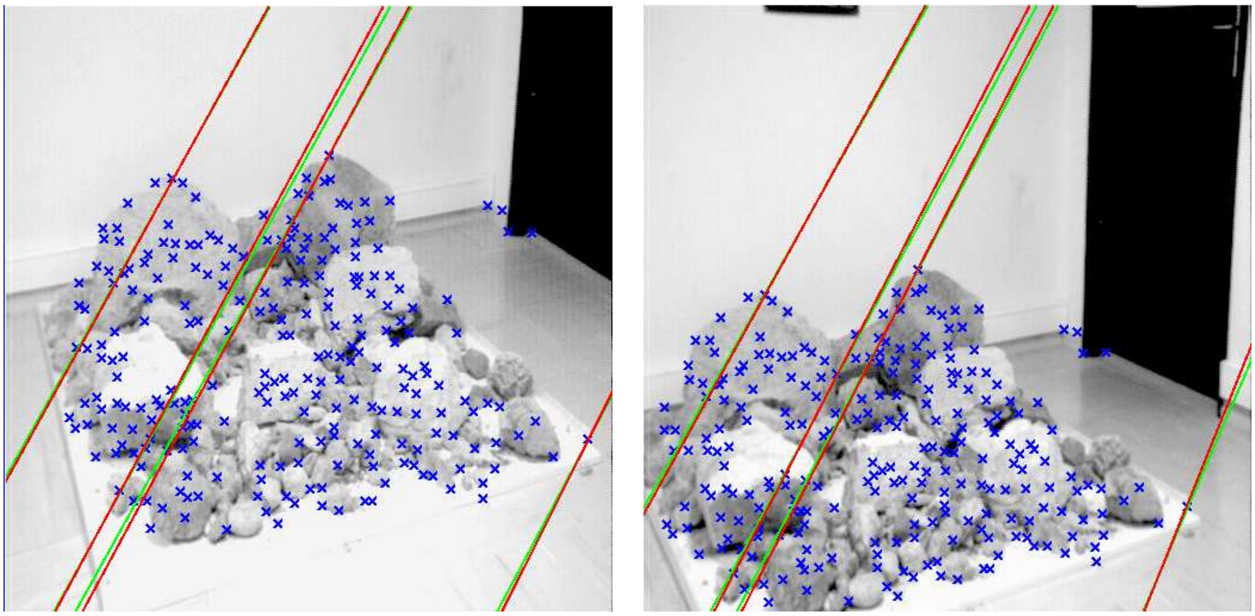

Figure 9: Comparison between the Epipolar geometry estimated through classical stereo calibration (shown in Red/Dark lines) and that estimated with the nonlinear method (shown in Green/Grey lines)

appropriate to measure the difference between two fundamental matrix. We will describe another one in the next paragraph. The third and fourth columns show the positions of the two epipoles. The fifth column gives the root of the mean of squared distances between points and their epipolar lines. We can see that even with **Calib**, the RMS is as high as 1 pixel. There are two possibilities: either the stereo system is not very well calibrated, or the points are not well localized; and we think the latter is the major reason because the corner detector we use only extracts points within pixel precision. The last column shows the approximate CPU time in seconds when the program is run on a Sparc 20 workstation. **Nonlinear**, **gradient** and **reproj**

Table 1: Comparison of different methods for estimating the fundamental matrix

Method	$\Delta\mathbf{F}$	$\mathbf{e}$		$\mathbf{e}'$		RMS	CPU
Calib.		5138.18	−8875.85	1642.02	−2528.91	0.99	
linear	5.85%	304.018	124.039	256.219	230.306	3.40	0.13s
normal.	7.20%	−3920.6	7678.71	8489.07	−15393.5	0.89	0.15s
nonlinear	0.92%	8135.03	−14048.3	1896.19	−2917.11	0.87	0.38s
gradient	0.92%	8166.05	−14104.1	1897.80	−2920.12	0.87	0.40s
M-estim.	0.12%	4528.94	−7516.3	1581.19	−2313.72	0.87	1.05s
reproj.	0.92%	8165.05	−14102.3	1897.74	−2920.01	0.87	19.1s
LMedS	0.13%	3919.12	−6413.1	1500.21	−2159.65	0.75	2.40s

give essentially the same result (but the latter is much more time consuming). The M-estimator and LMedS techniques give the best results. This is because the influence of poorly localized points has been reduced in M-estimator or they are simply discarded in LMedS. Actually, LMedS has detected five matches as outliers, which are 226, 94, 17, 78 and 100. Of course, these two methods are more time consuming than the nonlinear method.

4.1 A Measure of Comparison Between Fundamental Matrices

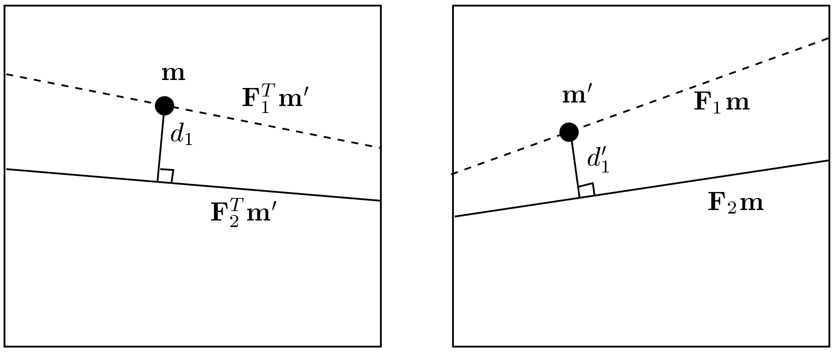

Figure 10: Definition of the difference between two fundamental matrices in terms of image distances

From the above discussion, the Frobenius norm of the difference between two normalized fundamental matrices is clearly not an appropriate measure of comparison. In the following, we describe a measure proposed by Stéphane Laveau from INRIA Sophia-Antipolis, which we think characterizes well the difference between two fundamental matrices. Let the two given fundamental matrices be $\mathbf{F}_1$ and $\mathbf{F}_2$. The measure is computed as follows (see Fig. 10):

Step 1: Choose *randomly* a point $\mathbf{m}$ in the first image.

Step 2: Draw the epipolar line of $\mathbf{m}$ in the second image using $\mathbf{F}_1$. The line is shown as a dashed line, and is defined by $\mathbf{F}_1\mathbf{m}$.

Step 3: If the epipolar line does not intersect the second image, go to **Step 1**.

Step 4: Choose *randomly* a point $\mathbf{m}'$ on the epipolar line. Note that $\mathbf{m}$ and $\mathbf{m}'$ correspond to each other exactly with respect to $\mathbf{F}_1$.

Step 5: Draw the epipolar line of $\mathbf{m}$ in the second image using $\mathbf{F}_2$, i.e., $\mathbf{F}_2\mathbf{m}$, and compute the distance, noted by d_1', between point $\mathbf{m}'$ and line $\mathbf{F}_2\mathbf{m}$.

Step 6: Draw the epipolar line of $\mathbf{m}'$ in the first image using $\mathbf{F}_2$, i.e., $\mathbf{F}_2^T\mathbf{m}'$, and compute the distance, noted by d_1, between point $\mathbf{m}$ and line $\mathbf{F}_2^T\mathbf{m}'$.

Step 7: Conduct the same procedure from **Step 2** through **Step 6**, but reversing the roles of $\mathbf{F}_1$ and $\mathbf{F}_2$, and compute d_2 and d_2'.

Step 8: Repeat N times **Step 1** through **Step 7**.

Step 9: Compute the average distance of d's, which is the measure of difference between the two fundamental matrices.

In this procedure, a random number generator based on uniform distribution is used. The two fundamental matrices plays a symmetric role. The two images plays a symmetric role too, although it is not at first sight. The reason is that $\mathbf{m}$ and $\mathbf{m}'$ are chosen randomly and the epipolar lines are symmetric (line $\mathbf{F}_1^T\mathbf{m}'$ goes through $\mathbf{m}$). Clearly, the measure computed as above, *in pixels*, is physically meaningful, because it is defined in the image space in which we observe the surrounding environment. Furthermore, when N tends to infinity, we sample uniformly the whole 3D space visible from the given epipolar geometry. If the image resolution is 512×512 and if we consider a pixel resolution, then the visible 3D space can be approximately sampled by 512^3 points. In our experiment, we set $N = 50000$. Using this method, we can compute the distance between each pair of fundamental matrices, and we obtain a symmetric matrix. The result is shown in Table 2, where only the upper triangle is displayed (because of symmetry). We arrive at the following conclusions:

- The linear method is very bad.
- The linear method with prior data normalization gives quite a reasonable result.
- The nonlinear method based on point-line distances and that based on gradient-weighted epipolar errors give very similar results to those obtained based on minimization of distances between observed points and reprojected ones. The latter should be avoided because it is too time consuming.
- M-estimators or the LMedS method give still better results because they try to limit or eliminate the effect of poorly localized points. The epipolar geometry estimated by LMedS is closer to the one computed through stereo calibration.

The LMedS method should be definitely used if the given set of matches contain false matches.

Table 2: Distances between the fundamental matrices estimated by different techniques

	linear	normal.	nonlinear	gradient	M-estim.	reproj.	LMedS
Calib.	**116.4**	**5.97**	**2.66**	**2.66**	**2.27**	**2.66**	**1.33**
linear		117.29	115.97	116.40	115.51	116.25	115.91
normal.			4.13	4.12	5.27	4.11	5.89
nonlinear				0.01	1.19	0.01	1.86
gradient					1.19	0.00	1.86
M-estim.						1.20	1.03
reproj.							1.88

4.2 Epipolar Band

Due to space limitation, the result on the uncertainty of the fundamental matrix is not shown here, and can be found in (Csurka et al. 1996), together with its use in computing the uncertainty of the projective reconstruction and in improving the self-calibration based on Kruppa equations. We show in this section how to use the uncertainty to define the epipolar band for matching.

We only consider the epipolar lines in the second image (the same can be done for the first). For a given point $\mathbf{m}_0 = [u_0, v_0]^T$ in the first image together with its covariance matrix $\mathbf{\Lambda}_{\mathbf{m}_0} = \left[\begin{smallmatrix} \sigma_{uu} & \sigma_{uv} \\ \sigma_{uv} & \sigma_{vv} \end{smallmatrix}\right]$, its epipolar line in the second image is given by $\mathbf{l}'_0 = \mathbf{F}\widetilde{\mathbf{m}}_0$. From (34), the covariance matrix of $\mathbf{l}'$ is computed by

$$\mathbf{\Lambda}_{\mathbf{l}'_0} = \frac{\partial \mathbf{l}'_0}{\partial \mathbf{F}} \mathbf{\Lambda}_{\mathbf{F}} \frac{\partial \mathbf{l}'_0}{\partial \mathbf{F}}^T + \mathbf{F} \left[\begin{smallmatrix} \mathbf{\Lambda}_{\mathbf{m}_0} & \mathbf{0}_2 \\ \mathbf{0}_2^T & 0 \end{smallmatrix}\right] \mathbf{F}^T , \tag{53}$$

where $\mathbf{F}$ in the first term of the right hand is treated as a 9-vector, and $\mathbf{0}_2 = [0, 0]^T$.

Any point $\mathbf{m}' = [u', v']^T$ on the epipolar line $\mathbf{l}'_0 \equiv [l'_1, l'_2, l'_3]^T$ must satisfy $\widetilde{\mathbf{m}}'^T \mathbf{l}'_0 = \mathbf{l}'^T_0 \widetilde{\mathbf{m}}' = l'_1 u' + l'_2 v' + l'_3 = 0$ (we see the *duality* between points and lines). The vector $\mathbf{l}'_0$ is defined up to a scale factor. It is a projective point in the dual space of the image plane, and is the dual of the epipolar line. We consider the vector of parameters $\mathbf{x}_0 = (x_0, y_0) = (l'_1/l'_3, l'_2/l'_3)^T$ (if $l'_3 = 0$ we can choose $(l'_1/l'_2, l'_3/l'_2)$ or $(l'_2/l'_1, l'_3/l'_1)$). The covariance matrix of $\mathbf{x}_0$ is computed in the same way as (34): $\mathbf{C} = (\partial \mathbf{l}'_0/\partial \mathbf{x}_0)\mathbf{\Lambda}_{\mathbf{l}'_0}(\partial \mathbf{l}'_0/\partial \mathbf{x}_0)^T$. The uncertainty of $\mathbf{x}_0$ can be represented in the usual way by an ellipse $\mathcal{C}$ in the dual space (denoted by $\mathbf{x}$) of the image plane:

$$(\mathbf{x} - \mathbf{x}_0)^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{x}_0) = k^2 , \tag{54}$$

where k is a confidence factor determined by the χ^2 distribution of 2 degrees of freedom. The probability that $\mathbf{x}$ appears at the interior of the ellipse defined by (54) is equal to $P_{\chi^2}(k, 2)$. Equation (54) can be rewritten in projective form as

$$\widetilde{\mathbf{x}}^T \mathbf{A} \widetilde{\mathbf{x}} = 0 \qquad \text{with} \quad \mathbf{A} = \begin{bmatrix} \mathbf{C}^{-1} & -\mathbf{C}^{-1}\mathbf{x}_0 \\ -\mathbf{x}_0^T \mathbf{C}^{-T} & \mathbf{x}_0^T \mathbf{C}^{-1} \mathbf{x}_0 - k^2 \end{bmatrix} .$$

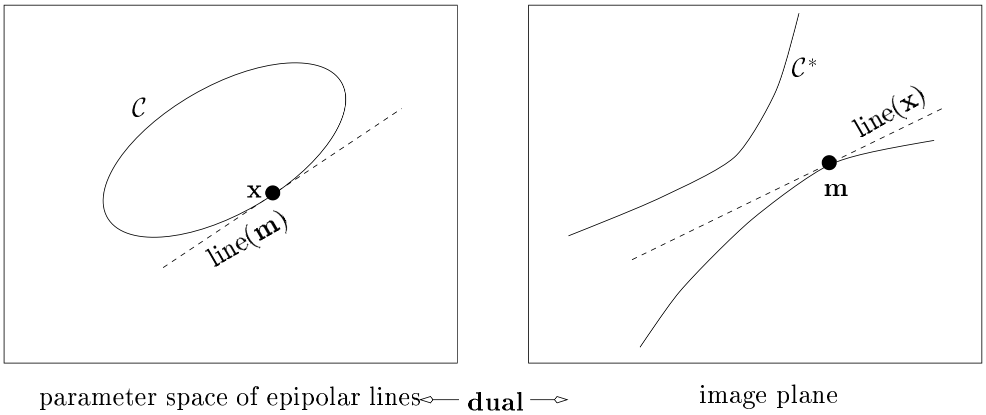

Figure 11: Duality between the image plane and the parameter space of the epipolar lines

The dual of this ellipse, denoted by $\mathcal{C}^*$, defines a conic in the image plane. It is given by

$$\widetilde{\mathbf{m}}^T \mathbf{A}^* \widetilde{\mathbf{m}} = 0 \tag{55}$$

where $\mathbf{A}^*$ is the adjoint of matrix $\mathbf{A}$ (i.e., $\mathbf{A}^*\mathbf{A} = \det(\mathbf{A})\,\mathbf{I}$). Because of the duality between the parameter space $\mathbf{x}$ and the image plane $\mathbf{m}$ (see Fig. 11), for a point $\mathbf{x}$ on $\mathcal{C}$, it defines an epipolar line in the image plane, line($\mathbf{x}$), which is tangent to conic $\mathcal{C}^*$ at a point $\mathbf{m}$, while the latter defines a line in the parameter space, line($\mathbf{m}$), which is tangent to $\mathcal{C}$ at $\mathbf{x}$. It can be shown (Csurka 1996) that, for a point in the interior of ellipse $\mathcal{C}$, the corresponding epipolar line lies outside of conic $\mathcal{C}^*$ (i.e., it does not cut the conic). Therefore, for a given k, the outside of this conic defines the region in which the epipolar line should lie with probability $P_{\chi^2}(k, 2)$. We call this region the *epipolar band*. For a given point in one image, its match should be searched in this region. Although, theoretically, the uncertainty conic defining the epipolar band could be an ellipse or parabola, it is always an hyperbola in practice (except when $\mathbf{\Lambda}_{\mathbf{F}}$ is extremely huge).

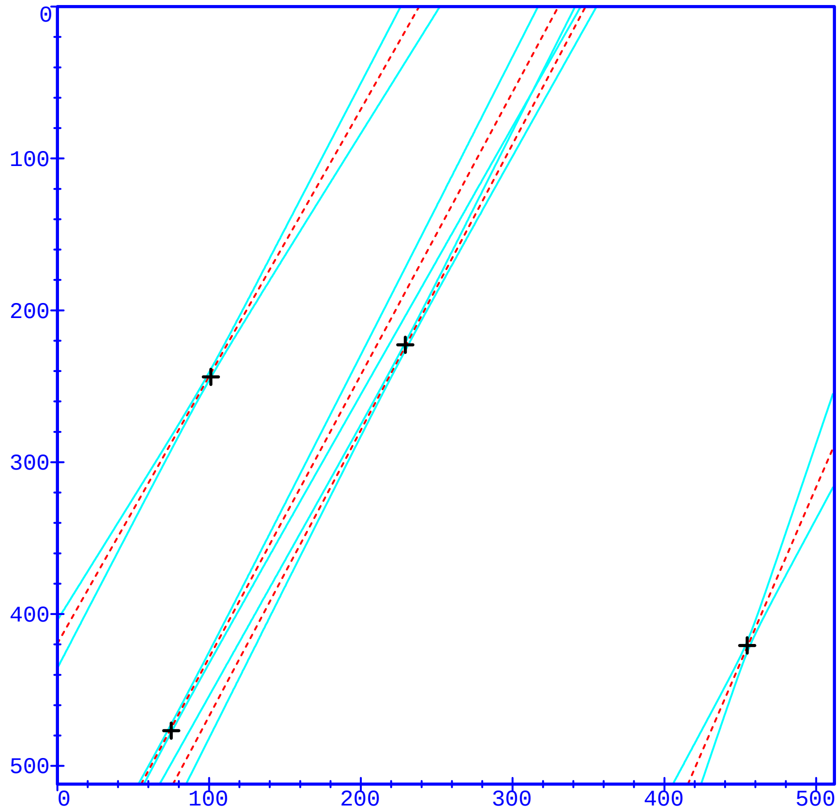

Figure 12: Epipolar bands for several point matches

We have estimated the uncertainty of the fundamental matrix for the image pair shown in Fig. 4. In Fig. 12, we show the epipolar bands of matches 1, 220, 0 and 183 in the second images, computed as described above. The displayed hyperbolas correspond to a probability of 70% ($k = 2.41$) with image point uncertainty of $\sigma_{uu} = \sigma_{vv} = 0.5^2$ and $\sigma_{uv} = 0$. We have also shown in Fig. 12 the epipolar lines drawn in dashed lines and the matched points indicated in $+$. An interesting thing is that the matched points are located in the area where the two sections of hyperbolas are closest to each other. This suggests that the covariance matrix of the fundamental matrix actually captures, to some extent, the matching information (disparity in stereo terminology). Such areas should be first examined in searching for point matches. This may, however, not be true if a significant depth discontinuity presents in the scene and if the point matches used in computing the fundamental matrix do not represent sufficiently enough the depth variation.

5 Discussion

In this paper, we have reviewed a number of techniques for estimating the epipolar geometry between two images. Point matches are assumed to be given, but some of them may have been incorrectly paired. How to establish point matches is the topic of the paper (Zhang, Deriche, Faugeras and Luong 1995).

5.1 Summary

For two uncalibrated images under full perspective projection, at least 7 point matches are necessary to determine the epipolar geometry. When only 7 matches are available, there are possibly three solutions, which can be obtained by solving a cubic equation. If more data are available, then the solution is in general unique and several linear techniques have been developed. The linear techniques are usually sensitive to noise and not very stable, because they ignore the constraints on the nine coefficients of the fundamental matrix and the criterion they are minimizing is not physically meaningful. The results, however, can be considerably improved by first normalizing the data points, instead of using pixel coordinates directly, such that their new coordinates are on the average equal to unity. Even better results can be obtained under nonlinear optimization framework by

- using an appropriate parameterization of fundamental matrix to take into account explicitly the rank-2 constraint, and
- minimizing a physically meaningful criterion.

Three choices are available for the latter: the distances between points and their corresponding epipolar lines, the gradient-weighted epipolar errors, and the distances between points and the reprojections of their corresponding points reconstructed in space. Experiments show that the results given by the optimization based on the first criterion are slightly worse than the last two which give essentially the same results. However, the third is much more time consuming, and is therefore not recommended, although it is statistically optimal under certain conditions. One can, however, use it as the last step to refine the results obtained with the first or second technique. To summarize, we recommend the second criterion (gradient-weighted epipolar errors), which is actually a very good approximation to the third one.

Point matches are obtained by using some heuristic techniques such as correlation and relaxation, and they usually contain false matches. Also, due to the limited performance of a corner detector or low contrast of an image, a few points are possibly poorly localized. These outliers (sometimes even one) will severely affect the precision of the fundamental matrix if we directly apply the methods described above, which are all least-squares techniques. We have thus presented in detail two commonly used robust techniques: M-Estimators and Least Median of Squares (LMedS). M-estimators try to reduce the effect of outliers by replacing the squared residuals by another function of the residuals which is less increasing than square. They can be implemented as an iterated reweighted least-squares. Experiments show that they are robust to outliers due to bad localization, but not robust to false matches. This is because they depend tightly on the initial estimation of the fundamental matrix. The LMedS method solves a nonlinear minimization problem which yields the smallest value for the median of squared residuals computed for the entire data set. It turns out that this method is very robust to false matches as well as to outliers due to bad localization. Unfortunately, there is no straightforward formula for the LMedS estimator. It must be solved by a search in the space of possible estimates generated from the data. Since this space is too large, only a randomly chosen subset of data can be analyzed. We have proposed a regularly random selection method to improve the efficiency.

Since the data points are always corrupted by noise, one should model the uncertainty of the estimated fundamental matrix in order to exploit its underlying geometric information correctly and effectively. We have modeled the fundamental matrix as a random vector in its parameterization space and described methods to estimate the covariance matrix of this vector under the first order approximation. This uncertainty measure can be used to define the epipolar band for matching, as shown in Sect. 4.2. In (Csurka et al. 1996), we also show how it can be used to compute the uncertainty of the projective reconstruction and to improve the self-calibration based on Kruppa equations.

Techniques for projective reconstruction will be reviewed in Appendix A. Although we cannot obtain any metric information from a projective structure (measurements of lengths and angles do not make sense), it still contains rich information, such as coplanarity, collinearity, and ratios, which is sometimes sufficient for artificial systems, such as robots, to perform tasks such as navigation and object recognition.

5.2 Degenerate configurations

Up to now, we have only considered the situations where no ambiguity arises in interpreting a set of point matches (i.e., they determine a unique fundamental matrix), except for the case of 7 point matches where three solutions may exist. Sometimes, however, even with a large set of point matches, there exist many solutions for the fundamental matrix which explain the data equally well, and we call such situations degenerate for the determination of the fundamental matrix.

Maybank (Maybank 1992) has thoroughly studied the degenerate configurations:

- 3D points lie on a quadric surface passing through the two optical centers (called the critical surface, or *maybank quadric* by Longuet-Higgins). We may have three different fundamental matrices compatible with the data. The two sets of image points are related by a quadratic transformation:

$$\mathbf{m}' = \mathbf{F}_1\mathbf{m} \times \mathbf{F}_2\mathbf{m} \ ,$$

  where $\mathbf{F}_1$ and $\mathbf{F}_2$ are two of the fundamental matrices.

- The two sets of image points are related by a homography:

$$\widetilde{\mathbf{m}}' = \rho\mathbf{H}\widetilde{\mathbf{m}} \ ,$$

  where ρ is an arbitrary non-zero scalar, and $\mathbf{H}$ is a 3×3 matrix defined up to a scale factor. This is a degenerate case of the previous situation. It arises when 3D points lie on a plane or when the camera undergoes a pure rotation around the optical center (equivalent to the case when all points lie on a plane at infinity).

- 3D points are in even more special position, for example on a line.

The stability of the fundamental matrix related to the degenerate configurations is analyzed in (Luong and Faugeras 1996). A technique which automatically detects the degeneracy based on χ^2 test when the noise level of the data points is known is reported in (Torr, Zisserman and Maybank 1995, Torr, Zisserman and Maybank 1996).

5.3 Affine cameras

So far, we have only considered images under perspective projection, which is a nonlinear mapping from 3D space to 2D. This makes many vision problems difficult to solve, and more importantly, they can become ill-conditioned when the perspective effects are small. Sometimes, if certain conditions are satisfied, for example, when the camera field of view is small and the object size is small enough with respect to the distance from the camera to the object, the projection can be approximated by a linear mapping (Aloimonos 1990). The affine camera introduced in (Mundy and Zisserman 1992) is a generation of the orthographic and weak perspective models. Its projection matrix has the following special form:

$$\mathbf{P}_A = \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} \\ P_{21} & P_{22} & P_{23} & P_{24} \\ 0 & 0 & 0 & P_{34} \end{bmatrix}$$

defined up to a scale factor. The epipolar constraint (5) is still valid, but the fundamental matrix (6) will be of the following simple form (Xu and Zhang 1996):

$$\mathbf{F}_A = \begin{bmatrix} 0 & 0 & a_{13} \\ 0 & 0 & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \ .$$

This is known as the *affine fundamental matrix* (Zisserman 1992, Shapiro, Zisserman and Brady 1994). Thus, the epipolar equation is linear in the image coordinates under affine cameras, and the determination of the epipolar geometry is much easier. This has been thoroughly studied by Oxford group (Shapiro 1993, Shapiro et al. 1994) (see also (Xu and Zhang 1996)), and thus is not addressed here. A software called `AffineF` is available from my Web home page.

5.4 Cameras with Lens Distortion

With the current formulation of the epipolar geometry (under either full perspective or affine projection), the homogeneous coordinates of a 3D point and those of the image point are related by a 3×4 matrix. That is, the lens distortion is not addressed. This statement does not imply, though, that lens distortion has never been accounted for in the previous work. Indeed, distortion has usually been corrected off-line using classical methods by observing for example straight lines, if it is not weak enough to be neglected. A preliminary investigation has been conducted (Zhang 1996b), which considers lens distortion as an integral part of a camera. In this case, for a point in one image, its corresponding point does not lie on a line anymore. As a matter of fact, it lies on the so-called *epipolar curve*. Preliminary results show that the distortion can be corrected on-line if cameras have a strong lens distortion. More work still needs to be done to understand better the epipolar geometry with lens distortion.

5.5 Multiple cameras

The study of the epipolar geometry is naturally extended to more images. When 3 images are considered, trilinear constraints exist between point/line correspondences (Spetsakis and Aloimonos 1989). "Trilinear" means that the constraints are linear in the point/line coordinates of *each image*, and the epipolar constraint (5) is a bilinear relation. The trilinear constraints have been rediscovered in (Shashua 1994b) in the context of uncalibrated images. Similar to the fundamental matrix for two images, the constraints between three images can be described by a $3 \times 3 \times 3$ matrix defined up to a scale factor (Spetsakis and Aloimonos 1989, Hartley 1994). There exist at most 4 linear independent constraints in the elements of the above matrix, and 7 point matches are required to have a linear solution (Shashua 1994b). However, the 27 elements are not algebraically independent. There are only 18 parameters to describe the geometry between three uncalibrated images (Faugeras and Robert 1994), and we have three algebraically independent constraints. Therefore, we need at least 6 point matches to determine the geometry of three images (Quan 1995).

When more images are considered, quadrilinear relations arising when four-tuples of images are considered, which are, however, algebraically dependent of the trilinear and bilinear ones (Faugeras and Mourrain 1995). That is, they do not bring in any new information. Recently, quite a lot of efforts have been directed towards the study of the geometry of N images (see (Luong and Viéville 1994, Carlsson 1994, Triggs 1995, Weinshall, Werman and Shashua 1995, Viéville, Faugeras and Luong 1996, Laveau 96) to name a few). A complete review of the work on multiple cameras is beyond the scope of this paper.

A Projective Reconstruction

We show in this section how to estimate the position of a point in space, given its projections in two images whose epipolar geometry is known. The problem is known as *3D reconstruction* in general, and *triangulation* in particular. In the calibrated case, the relative position (i.e., the rotation and translation) of the two cameras is known, and 3D reconstruction has already been extensively studied in stereo (Ayache 1991). In the uncalibrated case, like the one considered here, we assume that the fundamental matrix between the two images is known (e.g., computed with the methods described in Sect. 3), and we say that they are *weakly calibrated*.

A.1 Projective Structure from Two Uncalibrated Images

In the calibrated case, a 3D structure can be recovered from two images only up to a rigid transformation and an unknown scale factor (this transformation is also known as a *similarity*), because we can choose an arbitrary coordinate system as a world coordinate system (although one usually chooses it to coincide with one of the camera coordinate systems). Similarly, in the uncalibrated case, a 3D structure can only be performed up to a projective transformation of the 3D space (Maybank 1992, Faugeras 1992, Hartley et al. 1992, Faugeras 1995).

At this point, we have to introduce a few notations from Projective Geometry (a good introduction can be found in the appendix of (Mundy and Zisserman 1992) or (Faugeras 1995)). For a 3D point $\mathtt{M} = [X, Y, Z]^T$, its homogeneous coordinates are $\widetilde{\mathbf{x}} = [U, V, W, S]^T = \lambda\widetilde{\mathtt{M}}$ where λ is any nonzero scalar and $\widetilde{\mathtt{M}} = [X, Y, Z, 1]^T$. This implies: $U/S = X$, $V/S = Y$, $W/S = Z$. If we include the possibility that $S = 0$, then $\widetilde{\mathbf{x}} = [U, V, W, S]^T$ are called the *projective coordinates* of the 3D point $\mathtt{M}$, which are not all equal to zero and defined up to a scale factor. Therefore, $\widetilde{\mathbf{x}}$ and $\lambda\widetilde{\mathbf{x}}$ ($\lambda \neq 0$) represent the same projective point. When $S \neq 0$, $\widetilde{\mathbf{x}} = S\widetilde{\mathtt{M}}$. When $S = 0$, we say that the point is at infinity. A 4×4 nonsingular matrix $\mathbf{H}$ defines a linear transformation from one projective point to another, and is called the *projective transformation*. The matrix $\mathbf{H}$, of course, is also defined up to a nonzero scale factor, and we write

$$\rho\widetilde{\mathbf{y}} = \mathbf{H}\widetilde{\mathbf{x}} , \tag{56}$$

if $\widetilde{\mathbf{x}}$ is mapped to $\widetilde{\mathbf{y}}$ by $\mathbf{H}$. Here ρ is a nonzero scale factor.

Proposition 2. *Given two (perspective) images with unknown intrinsic parameters of a scene, the 3D structure of the scene can be reconstructed up to an unknown projective transformation as soon as the epipolar geometry (i.e., the fundamental matrix) between the two images is known.*

Assume that the true camera projection matrices are $\mathbf{P}$ and $\mathbf{P}'$. From (6), we have the following relation

$$\mathbf{F} = [\mathbf{P}'\mathbf{p}^{\perp}]_{\times}\mathbf{P}'\mathbf{P}^{+} ,$$

where $\mathbf{F}$ is the known fundamental matrix. The 3D structure thus reconstructed is $\mathtt{M}$. The proposition says that the 3D structure $\mathbf{H}^{-1}\widetilde{\mathtt{M}}$, where $\mathbf{H}$ is any projective transformation of the 3D space, is still consistent with the observed image points and the fundamental matrix. Following the pinhole model, the camera projection matrices corresponding to the new structure $\mathbf{H}^{-1}\widetilde{\mathtt{M}}$ are

$$\widehat{\mathbf{P}} = \mathbf{PH} \quad \text{and} \quad \widehat{\mathbf{P}}' = \mathbf{P}'\mathbf{H} ,$$

respectively. In order to show the above proposition, we only need to prove

$$[\widehat{\mathbf{P}}'\widehat{\mathbf{p}}^{\perp}]_{\times}\widehat{\mathbf{P}}'\widehat{\mathbf{P}}^{+} = \lambda\mathbf{F} \equiv \lambda[\mathbf{P}'\mathbf{p}^{\perp}]_{\times}\mathbf{P}'\mathbf{P}^{+} , \tag{57}$$

where $\widehat{\mathbf{p}}^{\perp} = (\mathbf{I} - \widehat{\mathbf{P}}^{+}\widehat{\mathbf{P}})\widehat{\boldsymbol{\omega}}$ with $\widehat{\boldsymbol{\omega}}$ any 4-vector, and λ is a scalar since $\mathbf{F}$ is defined up to a scale factor. The above result has been known for several years. In (Xu and Zhang 1996), we provide a simple proof through pure linear algebra.

A.2 Computing Camera Projection Matrices

The projective reconstruction is very similar to the 3D reconstruction when cameras are calibrated. First, we need to compute the camera projection matrices from the fundamental matrix $\mathbf{F}$ with respect to a projective basis, which can be arbitrary because of Proposition 2.

A.2.1 Factorization Method

Let $\mathbf{F}$ be the fundamental matrix for the two cameras. There are an infinite number of projective bases which all satisfy the epipolar geometry. One possibility is to factor $\mathbf{F}$ as a product of an antisymmetric matrix $[\mathbf{e}']_{\times}$ ($\mathbf{e}'$ is in fact the epipole in the second image) and a matrix $\mathbf{M}$, i.e., $\mathbf{F} = [\mathbf{e}']_{\times}\mathbf{M}$. A canonical representation can then be used:

$$\mathbf{P} = [\mathbf{I}\ \mathbf{0}] \quad \text{and} \quad \mathbf{P}' = [\mathbf{M}\ \mathbf{e}'] .$$

It is easy to verify that the above $\mathbf{P}$ and $\mathbf{P}'$ do yield the fundamental matrix.

The factorization of $\mathbf{F}$ into $[\mathbf{e}']_{\times}\mathbf{M}$ is in general not unique, because if $\mathbf{M}$ is a solution then $\mathbf{M} + \mathbf{e}'\mathbf{v}^T$ is also a solution for any vector $\mathbf{v}$ (indeed, we have always $[\mathbf{e}']_{\times}\mathbf{e}'\mathbf{v}^T = \mathbf{0}$). One way to do the factorization is as follow (Luong and Viéville 1994). Since $\mathbf{F}^T\mathbf{e}' = \mathbf{0}$, the epipole in the second image is given by the eigenvector of matrix $\mathbf{FF}^T$ associated to the smallest eigenvalue. Once we have $\mathbf{e}'$, using the relation

$$\|\mathbf{v}\|^2\mathbf{I}_3 = \mathbf{v}\mathbf{v}^T - [\mathbf{v}]_{\times}^2 \qquad \forall\mathbf{v} ,$$

we have

$$\mathbf{F} = \frac{1}{\|\mathbf{e}'\|^2}(\mathbf{e}'\mathbf{e}'^T - [\mathbf{e}']_{\times}^2)\mathbf{F} = \frac{1}{\|\mathbf{e}'\|^2}\underbrace{\mathbf{e}'\mathbf{e}'^T\mathbf{F}}_{\mathbf{0}} + [\mathbf{e}']_{\times}\underbrace{\left(-\frac{[\mathbf{e}']_{\times}}{\|\mathbf{e}'\|^2}\mathbf{F}\right)}_{\mathbf{M}} .$$

The first term on the right hand is equal to 0 because $\mathbf{F}^T\mathbf{e}' = \mathbf{0}$. We can thus define the $\mathbf{M}$ matrix as

$$\mathbf{M} = -\frac{1}{\|\mathbf{e}'\|^2}[\mathbf{e}']_\times \mathbf{F} \ .$$

This decomposition is used in (Beardsley et al. 1994). Numerically, better results of 3D reconstruction are obtained when the epipole $\mathbf{e}$ is normalized such that $\|\mathbf{e}\| = 1$.

A.2.2 Choosing a Projective Basis

Another possibility is to choose effectively five pairs of points, each of four points not being coplanar, between the two cameras as a projective basis. We can of course choose five corresponding points we have identified. However, the precision of the final projective reconstruction will depend heavily upon the precision of the pairs of points. In order to overcome this problem, we have chosen in (Zhang, Faugeras and Deriche 1995) the following solution. We first choose five arbitrary points in the first image, noted by $\mathbf{m}_i$ ($i = 1, \ldots, 5$). Although they could be chosen arbitrarily, they are chosen such that they are well distributed in the image to have a good numerical stability. For each point $\mathbf{m}_i$, its corresponding epipolar line in the second image is given by $\mathbf{l}'_i = \mathbf{F}\mathbf{m}_i$. We can now choose an arbitrary point on $\mathbf{l}'_i$ as $\mathbf{m}'_i$, the corresponding point of $\mathbf{m}_i$. Finally, we should verify that none of four points is coplanar, which can be easily done using the fundamental matrix (Faugeras 1992, credited to Roger Mohr). The advantage of this method is that the five pairs of points satisfy exactly the epipolar constraint.

Once we have five pairs of points $(\mathbf{m}_i, \mathbf{m}'_i)$, ($i = 1, \ldots, 5$), we can compute the camera projection matrices as described in (Faugeras 1992). Assigning the projective coordinates (somewhat arbitrarily) to the five reference points, we have five image points and space points in correspondence, which provides 10 constraints on each camera projection matrix, leaving only one unknown parameter. This unknown can then be solved using the known fundamental matrix.

A.3 Reconstruction Techniques

Now that the camera projection matrices of the two images with respect to a projective basis are available, we can reconstruct 3D structures *with respect to that projective basis* from point matches.

A.3.1 Linear Methods

Given a pair of points in correspondence: $\mathbf{m} = [u, v]^T$ and $\mathbf{m}' = [u', v']^T$. Let $\widetilde{\mathbf{x}} = [x, y, z, t]^T$ be the corresponding 3D point in space with respect to the projective basis chosen before. Following the pinhole model, we have:

$$s\,[u,\ v,\ 1]^T = \mathbf{P}\,[x,\ y,\ z,\ t]^T \ , \tag{58}$$

$$s'\,[u',\ v',\ 1] = \mathbf{P}'\,[x,\ y,\ z,\ t]^T \ , \tag{59}$$

where s and s' are two arbitrary scalars. Let $\mathbf{p}_i$ and $\mathbf{p}'_i$ be the vectors corresponding to the i^{th} row of $\mathbf{P}$ and $\mathbf{P}'$, respectively. The two scalars can then be computed as: $s = \mathbf{p}_3^T\widetilde{\mathbf{x}}$, $s' = \mathbf{p}'^T_3\widetilde{\mathbf{x}}$. Eliminating s and s' from (58) and (59) yields the following equation:

$$\mathbf{A}\widetilde{\mathbf{x}} = \mathbf{0} \ , \tag{60}$$

where $\mathbf{A}$ is a 4×4 matrix given by

$$\mathbf{A} = [\mathbf{p}_1 - u\mathbf{p}_3,\ \mathbf{p}_2 - v\mathbf{p}_3,\ \mathbf{p}'_1 - u'\mathbf{p}'_3,\ \mathbf{p}'_2 - v'\mathbf{p}'_3]^T \ .$$

As the projective coordinates $\widetilde{\mathbf{x}}$ are defined up to a scale factor, we can impose $\|\widetilde{\mathbf{x}}\| = 1$, then the solution to (60) is well known (see also the description in Sect. 3.2.2) to be the eigenvector of the matrix $\mathbf{A}^T\mathbf{A}$ associated to the smallest eigenvalue.

If we assume that no point is at infinity, then we can impose $t = 1$, and the projective reconstruction can be done exactly in the same way as for the Euclidean reconstruction. The set of homogeneous equations, $\mathbf{A}\widetilde{\mathbf{x}} = \mathbf{0}$, is reduced to a set of 4 non-homogeneous equations in 3 unknowns (x, y, z). A linear least-squares technique can be used to solve this problem.

A.3.2 Iterative Linear Methods

The previous approach has the advantage of providing a closed-form solution, but it has the disadvantage that the criterion that is minimized does not have a good physical interpretation. Let us consider the first of the equations (60). In general, the point $\widetilde{\mathbf{x}}$ found will not satisfy this equation exactly; rather, there will be an error $\epsilon_1 = \mathbf{p}_1^T \widetilde{\mathbf{x}} - u\mathbf{p}_3^T \widetilde{\mathbf{x}}$. What we really want to minimize is the difference between the measured image coordinate u and the projection of $\widetilde{\mathbf{x}}$, which is given by $\mathbf{p}_1^T \widetilde{\mathbf{x}} / \mathbf{p}_3^T \widetilde{\mathbf{x}}$. That is, we want to minimize

$$\epsilon_1' = \mathbf{p}_1^T \widetilde{\mathbf{x}} / \mathbf{p}_3^T \widetilde{\mathbf{x}} - u = \epsilon_1 / \mathbf{p}_3^T \widetilde{\mathbf{x}} \ .$$

This means that if the equation had been weighted by the factor $1/w_1$ where $w_1 = \mathbf{p}_3^T \widetilde{\mathbf{x}}$, then the resulting error would have been precisely what we wanted to minimize. Similarly, the weight for the second equation of (60) would be $1/w_2 = 1/w_1$, while the weight for the third and fourth equation would be $1/w_3 = 1/w_4 = 1/\mathbf{p}_3'^T \widetilde{\mathbf{x}}$. Finally, the solution could be found by applying exactly the same method described in the last subsection (either eigenvector computation or linear least-squares).

Like the method for estimating the fundamental matrix described in Sect. 3.4, the problem is that the weights w_i depends themselves on the solution $\widetilde{\mathbf{x}}$. To overcome this difficulty, we apply an iterative linear method. We first assume that all $w_i = 1$ and run a linear algorithm to obtain an initial estimation of $\widetilde{\mathbf{x}}$. The weights w_i are then computed from this initial solution. The weighted linear least-squares is then run for an improved solution. This procedure can be repeated several times until convergence (either the solution or the weight does not change between successive iterations). Two iterations are usually sufficient.

A.3.3 Nonlinear Methods

As said in the last paragraph, the quantity we want to minimize is the error measured in the image plane between the observation and the projection of the reconstruction, that is

$$(u - \frac{\mathbf{p}_1^T \widetilde{\mathbf{x}}}{\mathbf{p}_3^T \widetilde{\mathbf{x}}})^2 + (v - \frac{\mathbf{p}_2^T \widetilde{\mathbf{x}}}{\mathbf{p}_3^T \widetilde{\mathbf{x}}})^2 + (u' - \frac{\mathbf{p}_1'^T \widetilde{\mathbf{x}}}{\mathbf{p}_3'^T \widetilde{\mathbf{x}}})^2 + (v' - \frac{\mathbf{p}_2'^T \widetilde{\mathbf{x}}}{\mathbf{p}_3'^T \widetilde{\mathbf{x}}})^2 \ .$$

However, there does not exist any closed-form solution, and we must use any standard iterative minimization technique, such as the Levenberg-Marquardt. The initial estimate of $\widetilde{\mathbf{x}}$ can be obtained by using any linear technique described before.

Hartley and Sturm (1994) reformulates the above criterion in terms of the distance between a point and its corresponding epipolar line defined by the ideal space point being sought. By parameterizing the pencil of epipolar lines in one image by a parameter t (which defines also the corresponding epipolar line in the other image by using the fundamental matrix), they are able to transform the minimization problem to the resolution of a polynomial of degree 6 in t. There may exist up to 6 real roots, and the global minimum can be found by evaluating the minimization function for each real root.

More projective reconstruction techniques can be found in (Hartley and Sturm 1994, Rothwell, Csurka and Faugeras 1995), but it seems to us that the iterative linear or the nonlinear techniques based on the image errors are the best that one can recommend.

B Approximate Estimation of Fundamental Matrix from a General Matrix

We first introduce the Frobenius norm of a matrix $\mathbf{A} = [a_{ij}]$ ($i = 1, \ldots, m$; $j = 1, \ldots, n$), which is defined by

$$\|\mathbf{A}\| = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2} \ . \tag{61}$$

It is easy to show that for all orthogonal matrices $\mathbf{U}$ and $\mathbf{V}$ of appropriate dimensions, we have

$$\|\mathbf{U}\mathbf{A}\mathbf{V}^T\| = \|\mathbf{A}\| \ .$$

Proposition 3. *We are given a* 3×3 *matrix* $\mathbf{F}$*, whose singular value decomposition (SVD) is*

$$\mathbf{F} = \mathbf{U}\mathbf{S}\mathbf{V}^T \ ,$$

where $\mathbf{S} = diag(\sigma_1, \sigma_2, \sigma_3)$ *and* σ_i $(i = 1, 2, 3)$ *are singular values satisfying* $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq 0$*. Let* $\hat{\mathbf{S}} = diag(\sigma_1, \sigma_2, 0)$*, then*

$$\hat{\mathbf{F}} = \mathbf{U}\hat{\mathbf{S}}\mathbf{V}^T$$

is the closest matrix to $\mathbf{F}$ *that has rank 2. Here, "closest" is quantified by the Frobenius norm of* $\mathbf{F} - \hat{\mathbf{F}}$*, i.e.,* $\|\mathbf{F} - \hat{\mathbf{F}}\|$*.*

Proof. We show this in two parts.

First, the Frobenius norm of $\mathbf{F} - \hat{\mathbf{F}}$ is given by

$$\|\mathbf{F} - \hat{\mathbf{F}}\| = \|\mathbf{U}^T(\mathbf{F} - \hat{\mathbf{F}})\mathbf{V}\| = \|\operatorname{diag}(0, 0, \sigma_3)\| = \sigma_3 \ .$$

Second, for some 3×3 matrix $\mathbf{G}$ of *rank 2*, we can always find an orthogonal vector $\mathbf{z}$ such that $\mathbf{G}\mathbf{z} = \mathbf{0}$, i.e., $\mathbf{z}$ is the null vector of matrix $\mathbf{G}$. Since

$$\mathbf{F}\mathbf{z} = \sum_{i=1}^{3} \sigma_i (\mathbf{v}_i^T \mathbf{z}) \mathbf{u}_i \ ,$$

where $\mathbf{u}_i$ and $\mathbf{v}_i$ are the i^{th} column vectors of $\mathbf{U}$ and $\mathbf{V}$, we have

$$\begin{aligned} \|\mathbf{F} - \mathbf{G}\|^2 &\geq \|(\mathbf{F} - \mathbf{G})\mathbf{z}\|^2 = \|\mathbf{F}\mathbf{z}\|^2 \\ &= \sum_{i=1}^{3} \sigma_i^2 (\mathbf{v}_i^T \mathbf{z})^2 \geq \sigma_3^2 \ . \end{aligned}$$

This implies that $\hat{\mathbf{F}}$ is indeed the closest to $\mathbf{F}$, which completes the proof. □

In the above derivation, we have used the following inequality which relates the Frobenius norm to the vector norm:

$$\|\mathbf{A}\| \geq \max_{\|\mathbf{z}\|=1} \|\mathbf{A}\mathbf{z}\| \geq \|\mathbf{A}\mathbf{z}\| \text{ with } \|\mathbf{z}\| = 1.$$

The reader is referred to (Golub and van Loan 1989) for more details.

C Image Coordinates and Numerical Conditioning of Linear Least-Squares

This section describes the relation between the numerical conditioning of linear least-squares problems and the image coordinates, based on the analysis given in (Hartley 1995).

Consider the method described in Sect. 3.2.2, which consists in finding the eigenvector of the 9×9 matrix $\mathbf{U}_n^T \mathbf{U}_n$ associated with the least eigenvalue (for simplicity, this vector is called the *least eigenvector* in the sequel). This matrix can be expressed as $\mathbf{U}_n^T \mathbf{U}_n = \mathbf{U}\mathbf{D}\mathbf{U}^T$, where $\mathbf{U}$ is orthogonal and $\mathbf{D}$ is diagonal whose diagonal entries λ_i $(i = 1, \ldots, 9)$ are assumed to be in non-increasing order. In this case, the least eigenvector of $\mathbf{U}_n^T \mathbf{U}_n$ is the last column of $\mathbf{U}$. The ratio λ_1 / λ_8, denoted by κ, is the *condition number* of the matrix $\mathbf{U}_n^T \mathbf{U}_n$ (because λ_9 is expected to be 0). This parameter is well known to be an important factor in the analysis of stability of linear problems (Golub and van Loan 1989). If κ is large, then very small changes to the data can cause large changes to the solution. The sensitivity of invariant subspaces is discussed in detail in (Golub and van Loan 1989, p.413).

The major reason for the poor condition of the matrix $\mathbf{U}_n^T \mathbf{U}_n \equiv \mathbf{X}$ is the lack of homogeneity in the image coordinates. In an image of dimension 200×200, a typical image point will be of the form $(100, 100, 1)$.

If both $\widetilde{\mathbf{m}}_i$ and $\widetilde{\mathbf{m}}'_i$ are of this form, then $\mathbf{u}_i$ will be of the form $[10^4, 10^4, 10^2, 10^4, 10^4, 10^2, 10^2, 10^2, 1]^T$. The contribution to the matrix $\mathbf{X}$ is of the form $\mathbf{u}_i\mathbf{u}_i^T$, which will contain entries ranging between 10^8 and 1. The diagonal entries of $\mathbf{X}$ will be of the form $[10^8, 10^8, 10^4, 10^8, 10^8, 10^4, 10^4, 10^4, 1]^T$. Summing over all point matches will result in a matrix $\mathbf{X}$ whose diagonal entries are approximately in this proportion. We denote by $\mathbf{X}_r$ the trailing $r \times r$ principal submatrix (that is the last r columns and rows) of $\mathbf{X}$, and by $\lambda_i(\mathbf{X}_r)$ its i^{th} largest eigenvalue. Thus $\mathbf{X}_9 = \mathbf{X} = \mathbf{U}_n^T\mathbf{U}_n$ and $\kappa = \lambda_1(\mathbf{X}_9)/\lambda_8(\mathbf{X}_9)$. First, we consider the eigenvalues of $\mathbf{X}_2$. Since the sum of the two eigenvalues is equal to the trace, we see that $\lambda_1(\mathbf{X}_2)+\lambda_2(\mathbf{X}_2) = \text{trace}(\mathbf{X}_2) = 10^4+1$. Since eigenvalues are non-negative, we know that $\lambda_1(\mathbf{X}_2) \leq 10^4 + 1$. From the *interlacing property* (Golub and van Loan 1989, p.411), we arrive that

$$\lambda_8(\mathbf{X}_9) \leq \lambda_7(\mathbf{X}_8) \leq \cdots \leq \lambda_1(\mathbf{X}_2) \leq 10^4 + 1 \ .$$

On the other hand, also from the interlacing property, we know that the largest eigenvalue of $\mathbf{X}$ is not less than the largest diagonal entry, i.e., $\lambda_1(\mathbf{X}_9) \geq 10^8$. Therefore, the ratio $\kappa = \lambda_1(\mathbf{X}_9)/\lambda_8(\mathbf{X}_9) \geq 10^8/(10^4+1)$. In fact, $\lambda_8(\mathbf{X}_9)$ will usually be much smaller than 10^4+1 and the condition number will be far greater. This analysis shows that *scaling the coordinates so that they are on the average equal to* unity *will improve the condition of the matrix* $\mathbf{U}_n^T\mathbf{U}_n$.

Now consider the effect of translation. A usual practice is to fix the origin of the image coordinates at the top left hand corner of the image, so that all the image coordinates are positive. In this case, *an improvement in the condition of the matrix may be achieved by translating the points so that the centroid of the points is at the origin.* Informally, if the first image coordinates (the u-coordinates) of a set of points are $\{101.5, 102.3, 98.7, \ldots\}$, then the significant values of the coordinates are obscured by the coordinate offset of 100. By translating by 100, these numbers are changed to $\{1.5, 2.3, -1.3, \ldots\}$. The significant values become now prominent.

Thus, the conditioning of the linear least-squares process will be considerably improved by translating and scaling the image coordinates, as described in Sect. 3.2.5.

Acknowledgment

The author gratefully acknowledges the contribution of Gabiella Csurka, Stéphane Laveau, Gang Xu (Ritsumeikan University, Japan), and Cyril Zeller. The comments from Tuan Luong and Andrew Zisserman have helped the author to improve the paper.

References

Aggarwal, J. and Nandhakumar, N.: 1988, On the computation of motion from sequences of images — a review, *Proc. IEEE* **76**(8), 917–935.

Aloimonos, J.: 1990, Perspective approximations, *Image and Vision Computing* **8**(3), 179–192.

Anderson, T.: 1958, *An Introduction to Multivariate Statistical Analysis*, John Wiley & Sons, Inc.

Ayache, N.: 1991, *Artificial Vision for Mobile Robots*, MIT Press.

Ayer, S., Schroeter, P. and Bigün, J.: 1994, Segmentation of moving objects by robust motion parameter-estimation over multiple frames, *in* J.-O. Eklundh (ed.), *Proceedings of the 3rd European Conference on Computer Vision*, Vol. 800-801 of *Lecture Notes in Computer Science*, Springer-Verlag, Stockholm, Sweden, pp. 316–327, Vol.II.

Beardsley, P., Zisserman, A. and Murray, D.: 1994, Navigation using affine structure from motion, *in* J.-O. Eklundh (ed.), *Proceedings of the 3rd European Conference on Computer Vision*, Vol. 2 of *Lecture Notes in Computer Science*, Springer-Verlag, Stockholm, Sweden, pp. 85–96.

Boufama, B. and Mohr, R.: 1995, Epipole and fundamental matrix estimation using the virtual parallax property, *Proceedings of the 5th International Conference on Computer Vision*, IEEE Computer Society Press, Boston, MA, pp. 1030–1036.

Carlsson, S.: 1994, Multiple image invariance using the double algebra, *in* J. L. Mundy, A. Zissermann and D. Forsyth (eds), *Applications of Invariance in Computer Vision*, Vol. 825 of *Lecture Notes in Computer Science*, Springer-Verlag, pp. 145–164.

Csurka, G.: 1996, *Modélisation projective des objets tridimensionnels en vision par ordinateur*, PhD thesis, University of Nice, Sophia-Antipolis, France.

Csurka, G., Zeller, C., Zhang, Z. and Faugeras, O.: 1996, Characterizing the uncertainty of the fundamental matrix, *CVGIP: Image Understanding.* to appear. Updated version of INRIA Research Report 2560, 1995.

Deriche, R., Zhang, Z., Luong, Q.-T. and Faugeras, O.: 1994, Robust recovery of the epipolar geometry for an uncalibrated stereo rig, *in* J.-O. Eklundh (ed.), *Proceedings of the 3rd European Conference on Computer Vision*, Vol. 800-801 of *Lecture Notes in Computer Science*, Springer Verlag, Stockholm, Sweden, pp. 567–576, Vol. 1.

Enciso, R.: 1995, *Auto-Calibration des Capteurs Visuels Actifs. Reconstruction 3D Active.*, PhD thesis, Université Paris XI Orsay.

Faugeras, O.: 1992, What can be seen in three dimensions with an uncalibrated stereo rig, *in* G. Sandini (ed.), *Proceedings of the 2nd European Conference on Computer Vision*, Vol. 588 of *Lecture Notes in Computer Science*, Springer-Verlag, Santa Margherita Ligure, Italy, pp. 563–578.

Faugeras, O.: 1993, *Three-Dimensional Computer Vision: a Geometric Viewpoint*, The MIT Press.

Faugeras, O.: 1995, Stratification of 3-D vision: projective, affine, and metric representations, *Journal of the Optical Society of America A* **12**(3), 465–484.

Faugeras, O., Luong, T. and Maybank, S.: 1992, Camera self-calibration: theory and experiments, *in* G. Sandini (ed.), *Proc 2nd ECCV*, Vol. 588 of *Lecture Notes in Computer Science*, Springer-Verlag, Santa Margherita Ligure, Italy, pp. 321–334.

Faugeras, O. and Lustman, F.: 1988, Motion and Structure from Motion in a piecewise planar environment, *International Journal of Pattern Recognition and Artificial Intelligence* **2**(3), 485–508.

Faugeras, O. and Mourrain, B.: 1995, On the geometry and algebra of the point and line correspondences between n images, *Proceedings of the 5th International Conference on Computer Vision*, IEEE Computer Society Press, Boston, MA, pp. 951–956.

Faugeras, O. and Robert, L.: 1994, What can two images tell us about a third one?, *in* J.-O. Eklundh (ed.), *Proceedings of the 3rd European Conference on Computer Vision*, Vol. 800-801 of *Lecture Notes in Computer Science*, Springer-Verlag, Stockholm, Sweden. also INRIA Technical report 2018.

Fischler, M. and Bolles, R.: 1981, Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography, *Communications of the ACM* **24**, 381–385.

Golub, G. and van Loan, C.: 1989, *Matrix Computations*, The John Hopkins University Press.

Haralick, R.: 1986, Computer vision theory: The lack thereof, *Computer Vision, Graphics, and Image Processing* **36**, 372–386.

Hartley, R.: 1993, Euclidean reconstruction from uncalibrated views, *in* J. Mundy and A. Zisserman (eds), *Applications of Invariance in Computer Vision*, Vol. 825 of *Lecture Notes in Computer Science*, Springer-Verlag, Berlin, pp. 237–256.

Hartley, R.: 1994, Projective reconstruction and invariants from multiple images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **16**(10), 1036–1040.

Hartley, R.: 1995, In defence of the 8-point algorithm, *Proceedings of the 5th International Conference on Computer Vision*, IEEE Computer Society Press, Boston, MA, pp. 1064–1070.

Hartley, R., Gupta, R. and Chang, T.: 1992, Stereo from uncalibrated cameras, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Urbana Champaign, IL, pp. 761–764.

Hartley, R. and Sturm, P.: 1994, Triangulation, *Proceedings of the ARPA Image Understanding Workshop*, Defense Advanced Research Projects Agency, Morgan Kaufmann Publishers, Inc., pp. 957–966.

Heeger, D. J. and Jepson, A. D.: 1992, Subspace Methods for Recovering Rigid Motion I: Algorithm and Implementation, *The International Journal of Computer Vision* **7**(2), 95–117.

Hesse, O.: 1863, Die cubische Gleichung, von welcher die Lösung des Problems der Homographie von M. Chasles abhängt, *J. reine angew. Math.* **62**, 188–192.

Huang, T. and Netravali, A.: 1994, Motion and structure from feature correspondences: A review, *Proc. IEEE* **82**(2), 252–268.

Huber, P.: 1981, *Robust Statistics*, John Wiley & Sons, New York.

Laveau, S.: 1996, *Géométrie d'un système de N caméras. Théorie. Estimation. Applications.*, PhD thesis, École Polytechnique.

Longuet-Higgins, H.: 1981, A computer algorithm for reconstructing a scene from two projections, *Nature* **293**, 133–135.

Luong, Q.-T.: 1992, *Matrice Fondamentale et Calibration Visuelle sur l'Environnement-Vers une plus grande autonomie des systèmes robotiques*, PhD thesis, Université de Paris-Sud, Centre d'Orsay.

Luong, Q.-T. and Faugeras, O. D.: 1996, The fundamental matrix: Theory, algorithms and stability analysis, *The International Journal of Computer Vision* **1**(17), 43–76.

Luong, Q.-T. and Viéville, T.: 1994, Canonic representations for the geometries of multiple projective views, *in* J.-O. Eklundh (ed.), *Proceedings of the 3rd European Conference on Computer Vision*, Vol. 800-801 of *Lecture Notes in Computer Science*, Springer-Verlag, Stockholm, Sweden, pp. 589–599, Vol. 1.

Maybank, S.: 1992, *Theory of reconstruction From Image Motion*, Springer-Verlag.

Maybank, S. J. and Faugeras, O. D.: 1992, A theory of self-calibration of a moving camera, *The International Journal of Computer Vision* **8**(2), 123–152.

Mohr, R., Boufama, B. and Brand, P.: 1993, Accurate projective reconstruction, *in* J. Mundy and A. Zisserman (eds), *Applications of Invariance in Computer Vision*, Vol. 825 of *Lecture Notes in Computer Science*, Springer-Verlag, Berlin, pp. 257–276.

Mohr, R., Veillon, F. and Quan, L.: 1993, Relative 3d reconstruction using multiple uncalibrated images, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 543–548.

More, J.: 1977, The levenberg-marquardt algorithm, implementation and theory, *in* G. A. Watson (ed.), *Numerical Analysis*, Lecture Notes in Mathematics 630, Springer-Verlag.

Mundy, J. L. and Zisserman, A. (eds): 1992, *Geometric Invariance in Computer Vision*, MIT Press.

Odobez, J.-M. and Bouthemy, P.: 1994, Robust multiresolution estimation of parametric motion models applied to complex scenes, *Publication interne 788*, IRISA-INRIA Rennes, France.

Olsen, S.: 1992, Epipolar line estimation, *Proceedings of the 2nd European Conference on Computer Vision*, Santa Margherita Ligure, Italy, pp. 307–311.

Ponce, J. and Genc, Y.: 1996, Epipolar geometry and linear subspace methods: A new approach to weak calibration, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, San Francisco, CA, pp. 776–781.

Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vetterling, W. T.: 1988, *Numerical Recipes in C*, Cambridge University Press.

Quan, L.: 1993, Affine stereo calibration for relative affine shape reconstruction, *Proceedings of the fourth British Machine Vision Conference*, Surrey, England, pp. 659–668.

Quan, L.: 1995, Invariants of six points and projective reconstruction from three uncalibrated images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **17**(1).

Rey, W. J.: 1983, *Introduction to Robust and Quasi-Robust Statistical Methods*, Springer, Berlin, Heidelberg.

Robert, L. and Faugeras, O.: 1993, Relative 3d positioning and 3d convex hull computation from a weakly calibrated stereo pair, *Proceedings of the 4th International Conference on Computer Vision*, IEEE Computer Society Press, Berlin, Germany, pp. 540–544. also INRIA Technical Report 2349.

Rothwell, C., Csurka, G. and Faugeras, O.: 1995, A comparison of projective reconstruction methods for pairs of views, *Proceedings of the 5th International Conference on Computer Vision*, IEEE Computer Society Press, Boston, MA, pp. 932–937.

Rousseeuw, P. and Leroy, A.: 1987, *Robust Regression and Outlier Detection*, John Wiley & Sons, New York.

Shapiro, L.: 1993, *Affine Analysis of Image Sequences*, PhD thesis, University of Oxford, Department of Engineering Science, Oxford, UK.

Shapiro, L. and Brady, M.: 1995, Rejecting outliers and estimating errors in an orthogonal-regression framework, *Phil. Trans. Royal Soc. of Lon. A* **350**, 407–439.

Shapiro, L., Zisserman, A. and Brady, M.: 1994, Motion from point matches using affine epipolar geometry, *in* J.-O. Eklundh (ed.), *Proceedings of the 3rd European Conference on Computer Vision*, Vol. II of *Lecture Notes in Computer Science*, Springer-Verlag, Stockholm, Sweden, pp. 73–84.

Shashua, A.: 1994a, Projective structure from uncalibrated images: structure from motion and recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **16**(8), 778–790.

Shashua, A.: 1994b, Trilinearity in visual recognition by alignment, *in* J.-O. Eklundh (ed.), *Proceedings of the 3rd European Conference on Computer Vision*, Vol. 800-801 of *Lecture Notes in Computer Science*, Springer-Verlag, Stockholm, Sweden, pp. 479–484.

Spetsakis, M. and Aloimonos, J.: 1989, A unified theory of structure from motion, *Technical Report CAR-TR-482*, Computer Vision Laboratory, University of Maryland.

Sturm, R.: 1869, Das problem der projektivität und seine anwendung auf die flächen zweiten grades, *Math. Ann.* **1**, 533–574.

Torr, P.: 1995, *Motion Segmentation and Outlier Detection*, PhD thesis, Department of Engineering Science, University of Oxford.

Torr, P., Beardsley, P. and Murray, D.: 1994, Robust vision, *British Machine Vision Conference*, University of York, UK, pp. 145–154.

Torr, P. and Murray, D.: 1993, Outlier detection and motion segmentation, *in* P. Schenker (ed.), *Sensor Fusion VI, SPIE Vol.2059*, Boston, pp. 432–443.

Torr, P., Zisserman, A. and Maybank, S.: 1995, Robust detection of degenerate configurations for the fundamental matrix, *Proceedings of the 5th International Conference on Computer Vision*, IEEE Computer Society Press, Boston, MA, pp. 1037–1042.

Torr, P., Zisserman, A. and Maybank, S.: 1996, Robust detection of degenerate configurations whilst estimating the fundamental matrix, *Technical Report OUEL 2090/96*, Oxford University, Dept. of Engineering Science.

Triggs, B.: 1995, Matching constraints and the joint image, *Proceedings of the 5th International Conference on Computer Vision*, IEEE Computer Society Press, Boston, MA, pp. 338–343.

Tsai, R. and Huang, T.: 1984, Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surface, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**(1), 13–26.

Viéville, T., Faugeras, O. D. and Luong, Q.-T.: 1996, Motion of points and lines in the uncalibrated case, *The International Journal of Computer Vision* **17**(1), 7–42.

Weinshall, D., Werman, M. and Shashua, A.: 1995, Shape tensors for efficient and learnable indexing, *IEEE Workshop on Representation of Visual Scenes*, IEEE, pp. 58–65.

Xu, G. and Zhang, Z.: 1996, *Epipolar Geometry in Stereo, Motion and Object Recognition: A Unified Approach*, Kluwer Academic Publishers.

Zeller, C.: 1996, *Calibration Projective Affine et Euclidienne en Vision par Ordinateur*, PhD thesis, École Polytechnique.

Zeller, C. and Faugeras, O.: 1994, Applications of non-metric vision to some visual guided tasks, *Proceedings of the International Conference on Pattern Recognition*, Computer Society Press, Jerusalem, Israel, pp. 132–136. A longer version in INRIA Tech Report RR2308.

Zhang, Z.: 1995, Motion and structure of four points from one motion of a stereo rig with unknown extrinsic parameters, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **17**(12), 1222–1227.

Zhang, Z.: 1996a, A new multistage approach to motion and structure estimation: From essential parameters to euclidean motion via fundamental matrix, *Research Report 2910*, INRIA Sophia-Antipolis, France.
URL: *ftp://krakatoa.inria.fr/pub/html/Papers/zhang:96.ps*

Zhang, Z.: 1996b, On the epipolar geometry between two images with lens distortion, *International Conference on Pattern Recognition*, Vol. I, pages 407–411, Vienna, Austria.
URL: *ftp://krakatoa.inria.fr/pub/html/Papers/zhang:96d.ps.gz*

Zhang, Z.: 1996c, Parameter estimation techniques: A tutorial with application to conic fitting, *Image and Vision Computing.* In Press. Also INRIA Research Report No.2676, October 1995.

Zhang, Z., Deriche, R., Faugeras, O. and Luong, Q.-T.: 1995, A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry, *Artificial Intelligence Journal* **78**, 87–119.

Zhang, Z., Deriche, R., Luong, Q.-T. and Faugeras, O.: 1994, A robust approach to image matching: Recovery of the epipolar geometry, *Proc. International Symposium of Young Investigators on Information\Computer\Control*, Beijing, China, pp. 7–28.

Zhang, Z. and Faugeras, O. D.: 1992, *3D Dynamic Scene Analysis: A Stereo Based Approach*, Springer, Berlin, Heidelberg.

Zhang, Z., Faugeras, O. and Deriche, R.: 1995, Calibrating a binocular stereo through projective reconstruction using both a calibration object and the environment, *in* R. Mohr and C. Wu (eds), *Proc. Europe-China Workshop on Geometrical modelling and Invariants for Computer Vision*, Xi'an, China, pp. 253–260.

Zhang, Z., Luong, Q.-T. and Faugeras, O.: 1996, Motion of an uncalibrated stereo rig: self-calibration and metric reconstruction, *IEEE Trans. Robotics and Automation* **12**(1), 103–113.

Zhuang, X., Wang, T. and Zhang, P.: 1992, A highly robust estimator through partially likelihood function modeling and its application in computer vision, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **14**(1), 19–34.

Zisserman, A.: 1992, Notes on geometric invariants in vision, BMVC92 Tutorial.