The Fundamental matrix: theory, algorithms, and stability analysis

Q.-T. Luong*and O.D. Faugeras
I.N.R.I.A.
2004 route des Lucioles, B.P. 93
06902 Sophia-Antipolis, France
luong,faugeras@sophia.inria.fr

Abstract

In this paper we analyze in some detail the geometry of a pair of cameras, i.e. a stereo rig. Contrarily to what has been done in the past and is still done currently, for example in stereo or motion analysis, we do not assume that the intrinsic parameters of the cameras are known (coordinates of the principal points, pixels aspect ratio and focal lengths). This is important for two reasons. First, it is more realistic in applications where these parameters may vary according to the task (active vision). Second, the general case considered here, captures all the relevant information that is necessary for establishing correspondences between two pairs of images. This information is fundamentally *projective* and is hidden in a confusing manner in the commonly used formalism of the Essential matrix introduced by Longuet-Higgins [40]. This paper clarifies the projective nature of the correspondence problem in stereo and shows that the epipolar geometry can be summarized in one 3×3 matrix of rank 2 which we propose to call the Fundamental matrix.

After this theoretical analysis, we embark on the task of estimating the Fundamental matrix from point correspondences, a task which is of practical importance. We analyze theoretically, and compare experimentally using synthetic and real data, several methods of estimation. The problem of the stability of the estimation is studied from two complementary viewpoints. First we show that there is an interesting relationship between the Fundamental matrix and three-dimensional planes which induce homographies between the images and create unstabilities in the estimation procedures. Second, we point to a deep relation between the unstability of the estimation procedure and the presence in the scene of so-called *critical surfaces* which have been studied in the context of motion analysis. Finally we conclude by stressing the fact that we believe that the Fundamental matrix will play a crucial role in future applications of three-dimensional Computer Vision by greatly increasing its versatility, robustness and hence applicability to real difficult problems.

1 Introduction

Inferring three-dimensional information from images taken from different viewpoints is a central problem in computer vision. However, as the measured data in images are just pixel coordinates, there are only two approaches that can be used in order to perform this task:

The first one is to establish a model which relates pixel coordinates to 3D coordinates, and to compute the parameters of such a model. This is done by camera calibration [73] [20], which typically computes the 3×4 projection matrices $\tilde{\mathbf{P}}$, which relate the image pixel coordinates to a world reference frame. The 11 parameters of this projection matrix account for the internal geometry of the camera, as well as its position and orientation in space with respect to a

*present address: EECS, University of California, Berkeley CA 94720, e-mail: qtluong@robotics.eecs.berkeley.edu

fixed reference frame. The knowledge of the internal geometry of a camera allows us to obtain directions in 3D space from pixel measurements, and thus the usual Euclidean concepts can be used: the relative positioning of cameras is described by a rigid displacement, and the world is described by metric quantities. However, it is not always possible to assume that cameras can be calibrated off-line, particularly when using active vision systems. Another drawback is that by doing so, many parameters have to be estimated in the case of a stereo rig, namely 11 + 11 which in our opinion is much more than what is really needed in most applications. Even an approach where cameras are just calibrated individually for their 5 internal parameters, and the rigid displacement between them is estimated subsequently would require at least the estimation of 15 parameters.

Thus a second approach is emerging [55], which consists in using projective information , whose non-metric nature allows to use cameras whose internal parameters are unknown, This approach requires only geometric information which relates the different viewpoints, thus a much more small number of parameters have to be estimated. They also lead to a deeper understanding of the fundamental elements in the geometry of two cameras, being very naturally related to the image formation process. The geometric relations between the two cameras are described in projective terms rather than in Euclidean ones. We will see in this paper that only 7 parameters are sufficient to describe the projective relation of two cameras. This information is entirely contained in a matrix called the *Fundamental matrix*, thus it is very important to develop precise techniques to compute it, and to study their stability with respect to various 3D point configurations and different camera displacements.

In spite of the fact that there has been some confusion between the Fundamental matrix and Longuet-Higgins' Essential matrix, it is now known that the fundamental matrix can be computed from pixel coordinates of corresponding points in uncalibrated images, which is the basic data we start from, in this paper. Methods to obtain such correspondences at a subpixel precision are now available, but are detailed elsewhere [42, 10, 10], since the emphasis of the paper is on geometric and algebraic relations which can be used to compute the fundamental matrix and to analyze its stability. Line correspondences are not sufficient with two views. Another approach is to use linear filters tuned to a range of orientations and scales. Jones and Malik [32] have shown that it is also possible in this framework to recover the location of epipolar lines.

In section 2, we clarify the concept of Fundamental matrix, and show its relation with the epipolar transformation and the Essential matrix, and propose some parameterizations for its computation. In section 3, we proceed to analyze several methods to compute the Fundamental matrix in the general case and show, using both large sets of simulations and real data, that our non-linear computation techniques provide significant improvement in the accuracy of the Fundamental matrix determination over linear techniques. In section 4 we examine the case of planes, and point out an important relation between the Fundamental matrix and homography matrices, which yield unstability, but allows some new specific algorithms to be applied. The stability of the computation is investigated in a more general framework where the influence of the camera motion is also considered in section 5. We end the paper by pointing to several applications of the Fundamental matrix in order to stress its importance in computer vision.

2 The Fundamental Matrix

2.1 Notations

In this paper we use boldface letters for vectors and matrixes i.e. $\mathbf{x}$. Transposition of vectors and matrices is indicated by ^{T}, i.e. $\mathbf{x}^T$. The line between two points M_1 and M_2 is represented by $\langle M_1, M_2 \rangle$. The cross-product of two three-dimensional vector $\mathbf{x}$ and $\mathbf{y}$ is represented by $\mathbf{x} \times \mathbf{y}$. The antisymmetric matrix such that $[\mathbf{v}]_\times \mathbf{x} = \mathbf{v} \times \mathbf{x}$ for all vectors $\mathbf{x}$ is noted $[\mathbf{v}]_\times$. We differenciate between the projective geometric objects themselves and their representations. For example, a point in the image plane will be denoted by m whereas one of its coordinate vectors

will be denoted by $\mathbf{m}$.

2.2 The projective model

The camera model which we consider is the pinhole model. In this model, the camera performs a perspective projection of an object point M onto a pixel m in the retinal plane through the optical center C. The optical axis is the line going through C and perpendicular to the retinal plane. It pierces that plane at point c. If we consider an orthonormal system of coordinates in the retinal plane, which we call *normalized coordinates*, centered at c, say (c, u, v) we can define a three-dimensional orthonormal system of coordinates centered at the optical center C with two axes of coordinates parallel to the retinal ones and the third one parallel to the optical axis (C, x, y, z). In these two systems of coordinates, the relationship between the coordinates of m, image of M is particularly simple:

$$u = \frac{x}{z} \quad v = \frac{y}{z}$$

It is nonlinear but if we write it using the homogeneous (projective) coordinates of m and M, it becomes linear:

$$\begin{bmatrix} U \\ V \\ S \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ T \end{bmatrix} \tag{1}$$

In this equation U, V and T are the projective coordinates of the pixel m and X, Y, Z, T are the projective coordinates of the point M. If M is not at infinity (i.e. $T \neq 0$), its Euclidean coordinates are $x = \frac{X}{T}$, $y = \frac{Y}{T}$, $z = \frac{Z}{T}$. Therefore, x, y and z can also be considered as the projective coordinates of the pixel m.

We write equation (1) in matrix form:

$$\mathbf{m} = \tilde{\mathbf{P}}\mathbf{M}$$

where $\tilde{\mathbf{P}}$ is the 3×4 matrix appearing in (1). Introducing projective coordinates induces a big simplification in the formulation of properties of cameras. It is one of the reasons why projective geometry is emerging as an attractive framework for computer vision [55]. In this paper, we assume that the reader is familiar with some elementary projective geometry. Such material can be found in classical mathematic textbooks such as [65, 6, 21], but also in the computer vision literature where it is presented in chapters of recent books [14, 34, 55], and articles [51, 33].

The main property of this camera model is thus that *the relationship between the world coordinates and the pixel coordinates is linear projective.* This property is independent of the choice of the coordinate systems in the retinal plane or in the three-dimensional space. Changing projective coordinates in the 3-D space is equivalent to multiplying matrix $\tilde{\mathbf{P}}$ to the right by a four by four matrix. Indeed, suppose that we have $\mathbf{M} = \mathbf{D}\mathbf{M}'$, then $\mathbf{m} = \tilde{\mathbf{P}}\mathbf{D}\mathbf{M}'$. A special case is the case of a rigid displacement represented by a rotation matrix $\mathbf{R}$ and a translation vector $\mathbf{t}$. We have then

$$\mathbf{D} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix}$$

Similarly, changing coordinates in the retinal plane is equivalent to multiplying matrix $\mathbf{P}$ to the left by a three by three matrix. Indeed, suppose that we have $\mathbf{m} = \mathbf{A}\mathbf{m}'$, then $\mathbf{m}' = \mathbf{A}^{-1}\tilde{\mathbf{P}}\mathbf{M}$. A special case is the case where the change of coordinates represents the change from the pixel coordinates to the normalized pixel coordinates [16, 14], accounting for the internal geometry of the camera. A pinhole camera can therefore be specified by a 3×4 matrix $\tilde{\mathbf{P}}$ which is defined up to a scale factor (it is a projective quantity) and is of rank 3 (an easy way is to see that this is the case for the matrix appearing in (1) and that this rank property is preserved by multiplication on the right with matrixes of rank 4 and multiplication on the left with matrixes of rank 3).

Since this will be used in what follows, let us now see how the optical center C can be recovered from the matrix $\tilde{\mathbf{P}}$. Let us decompose $\tilde{\mathbf{P}}$ as follows

$$\tilde{\mathbf{P}} = [\ \mathbf{P}\mathbf{p}\ ]$$

where $\mathbf{P}$ is a 3×3 matrix of rank 3 and $\mathbf{p}$ is a 3×1 vector. Let us assume without loss of generality that C is not at infinity and let $\tilde{\mathbf{C}} = [\mathbf{C}^T 1]^T$ be a projective representation of this point. $\mathbf{C}$ is its 3×1 Euclidean vector of coordinates and the component equal to 1 accounts for the fact that C is not at infinity. C satisfies the equation $\tilde{\mathbf{P}}\tilde{\mathbf{C}} = \mathbf{0}$ from which we conclude

$$\mathbf{C} = -\mathbf{P}^{-1}\mathbf{p}$$

2.3 The epipolar geometry and the Fundamental matrix

We now consider the case of two cameras looking at the same scene. The epipolar geometry is the basic constraint which arises from the existence of two viewpoints. Let us consider two images taken by linear projection from two different locations, as shown in figure 1. Let C be the optical center of the first camera, and let C' be the optical center of the second camera. The line $\langle C, C' \rangle$ projects to a point e in the first image $\mathcal{R}$, and to a point $e^{'}$ in the second image $\mathcal{R}'$. The points e, $e^{'}$ are the epipoles. The lines through e in the first image and the lines through $e^{'}$ in the second image are the epipolar lines. The epipolar constraint is well-known in stereovision: for each point m in the first retina, its corresponding point m' lies on its epipolar line l'_m. Similarly, for a given point m' in the second retina, its corresponding point m lies on its epipolar line $l_{m'}$. l'_m and $l_{m'}$ are called corresponding epipolar lines.

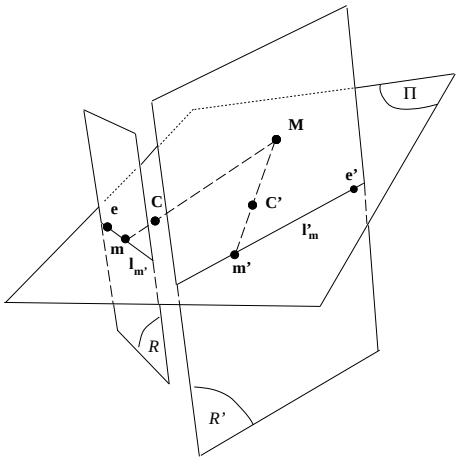

Figure 1: The epipolar geometry

The relationship between the retinal coordinates of a point m and its corresponding epipolar line l'_m is projective linear, because the relations between m and $\langle C, m \rangle$, and $\langle C, m \rangle$ and its projection l'_m are both projective linear. We call the 3×3 matrix $\mathbf{F}$ which describes this correspondence the *Fundamental matrix*. The importance of the Fundamental matrix has been neglected in the literature, as almost all the work on motion and stereo has been done under the assumption that intrinsic parameters are known. In that case, the Fundamental matrix reduces to an Essential matrix. But if one wants to proceed only from image measurements, the Fundamental matrix is the key concept, as it contains all the geometrical information relating two different images. One way to see it is to remember that the position along epipolar lines are related to the three-dimensional depth [61]. But if we do not have any knowledge about the scene geometry, we cannot infer such information.

Let us now express the epipolar constraint using the Fundamental matrix, in the case of uncalibrated cameras. For a given point m in the first image, the projective representation $\mathbf{l}'_m$ of its the epipolar line in the second image is given by

$$\mathbf{l}'_m = \mathbf{F}\mathbf{m}$$

Since the point $m^{'}$ corresponding to m belongs to the line l'_m by definition, it follows that:

$$\mathbf{m}^{'T}\mathbf{F}\mathbf{m} = 0 \tag{2}$$

Note that by reversing the role of the two images, the Fundamental matrix is changed to its transpose. Indeed, transposing equation (2), we obtain

$$\mathbf{m}^T\mathbf{F}^T\mathbf{m}' = 0$$

this shows that the epipolar line $l_{m'}$ of m' is represented by $\mathbf{F}^T\mathbf{m}$.

Just as in the one-camera case where we related the optical center to the perspective projection $\tilde{\mathbf{P}}$, in the two-cameras case, we can also relate the fundamental matrix $\mathbf{F}$ to the two perspective projection matrices $\tilde{\mathbf{P}}$ and $\tilde{\mathbf{P}}$. The epipole in the second image is the projection of the optical center of the first camera into the second camera, thus:

$$\mathbf{e}' = \tilde{\mathbf{P}}' \begin{bmatrix} -\mathbf{P}^{-1}\mathbf{p} \\ 1 \end{bmatrix} = \mathbf{p}' - \mathbf{P}'\mathbf{P}^{-1}\mathbf{p} \tag{3}$$

The epipolar line of a point m of the first retina is defined by the image from the second camera of two particular points of the optical ray $\langle C, M\rangle$: the optical center C (which is projected to the epipole e') and the point of infinity of $\langle C, M\rangle$. This point is projected to:

$$\tilde{\mathbf{P}}' \begin{bmatrix} \mathbf{P}^{-1}\mathbf{m} \\ 0 \end{bmatrix} = \mathbf{P}'\mathbf{P}^{-1}\mathbf{m}$$

The projective representation of the epipolar line l'_m is obtained by taking the cross-product of these two points, and it is seen again that this expression is linear in $\mathbf{m}$:

$$\mathbf{l}'_m = [\mathbf{p}' - \mathbf{P}'\mathbf{P}^{-1}\mathbf{p}] \times \mathbf{P}'\mathbf{P}^{-1}\mathbf{m} = \underbrace{[\mathbf{p}' - \mathbf{P}'\mathbf{P}^{-1}\mathbf{p}]_\times \mathbf{P}'\mathbf{P}^{-1}}_{\mathbf{F}}\mathbf{m} \tag{4}$$

2.4 Projective interpretation: Relation with the epipolar transformation

Let us enrich the idea of epipolar geometry and consider the one parameter family of planes going through $\langle C, C'\rangle$ as shown in figure 2. This family is a pencil of planes. Let Π be any plane containing $\langle C, C'\rangle$. Then Π projects to an epipolar line l in the first image and to an epipolar line $l^{'}$ in the second image. The correspondences $\Pi \overline{\wedge} l$ and $\Pi \overline{\wedge} l^{'}$ are homographies[1] between the two pencils of epipolar lines and the pencil of planes containing $\langle C, C'\rangle$. It follows that the correspondence $l \overline{\wedge} l^{'}$ is a homography, called the epipolar transformation. We relate it to the Fundamental matrix as follows.

We prove the following interesting property of the fundamental matrix. The fundamental matrix is such that

$$\mathbf{F}\mathbf{e} = \mathbf{F}^T\mathbf{e}' = \mathbf{0}$$

Indeed, the epipolar line of the epipole e is $\mathbf{F}\mathbf{e}$. Geometrically, this line l'_e is the image of the optical ray $\langle C, e\rangle$ in the second image. By construction this line is reduced to a point, e'. This implies that $\mathbf{l}'_e = \mathbf{F}\mathbf{e} = \mathbf{0}$. Same for the second epipole.

[1] It can be seen that by construction they preserve the cross-ratio.

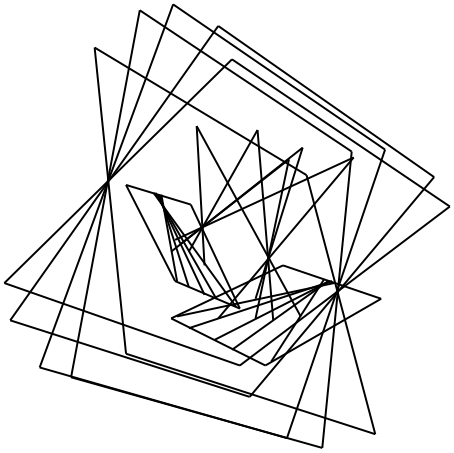

Figure 2: The epipolar pencils.

As a consequence of this, the rank of $\mathbf{F}$ is less than or equal to 2. In general, that rank is equal to 2. The case where it is equal to 1 is not possible since it can be seen that it implies that the line $\langle C, C' \rangle$ belongs to the two retinal planes and hence to their intersection. If we note $\mathbf{l}_i^T, i = 1, 2, 3$ the row vectors of $\mathbf{F}$ and $\mathbf{c}_i$ its column vectors, it means that we can write $\mathbf{e}$ (resp. $\mathbf{e}'$) as $\mathbf{l}_i \times \mathbf{l}_j$ (resp. $\mathbf{c}_i \times \mathbf{c}_j$) if $\mathbf{l}_i$ and $\mathbf{l}_j$ (resp. $\mathbf{c}_i$ and $\mathbf{c}_j$) are linearly independent.

We now find a parameterization of the pencils of epipolar lines such that the epipolar correspondence has a simple form. One solution, valid in the practical case where epipoles are at finite distance, and illustrated in figure 3, consists in intersecting each epipolar line in each retinal plane with the line at infinity l_∞ of that plane; these lines consist of retinal points for which the third projective component is zero. Their equation is $x_3 = 0$ and their projective representation is $(0, 0, 1)^T$. The epipolar transformation can then be expressed as a collineation between these two lines. If the epipolar line l goes through the point q, then its intersection with the line at infinity is $\mathbf{y}_\infty = (\mathbf{e} \times \mathbf{q}) \times (0, 0, 1)^T$, which can be written as $(1, \tau, 0)^T$, with:

$$\tau = \frac{q_2 - e_2}{q_1 - e_1} \tag{5}$$

Note that τ is the direction of the epipolar line l. Since the epipole is at finite distance, thus not on l_∞, it is an appropriate parameterization. If q' corresponds to q, then the epipolar line l' of the second image going through q' corresponds to l. It is parameterized by the point $\mathbf{y}'_\infty = (1, \tau', 0)^T$, with its projective parameter obtained by priming the quantities in (5). The epipolar transformation maps $\mathbf{y}_\infty$ to $\mathbf{y}'_\infty$, and thus is a homographic function of the projective parameters:

$$\tau \mapsto \tau' = \frac{a\tau + b}{c\tau + d} \tag{6}$$

Thus we can characterize the epipolar transformation by the four coordinates of the two epipoles $\mathbf{e}$ and $\mathbf{e}'$ and by the three coefficients of the homography. It follows that the epipolar transformation, like the Fundamental matrix depends on seven independent parameters.

Replacing τ and τ' by their values (5) in (6) and identifying the result with equation (2), we obtain expressions for the coefficients of $\mathbf{F}$ in terms of the parameters describing the epipoles and the homography:

$$\mathbf{F} = \begin{bmatrix} be_3e'_3 & ae_3e'_3 & -ae_2e'_3 - be_1e'_3 \\ -de'_3e_3 & -ce'_3e_3 & ce'_3e_2 + de'_3e_1 \\ de'_2e_3 - be_3e'_1 & ce'_2e_3 - ae_3e'_1 & -ce'_2e_2 - de'_2e_1 + ae_2e'_1 + be_1e'_1 \end{bmatrix} \tag{7}$$

The equations (7), yield the coefficients of the homography as functions of the Fundamental matrix:

$$a : b : c : d = F_{12} : F_{11} : -F_{22} : -F_{21} \tag{8}$$

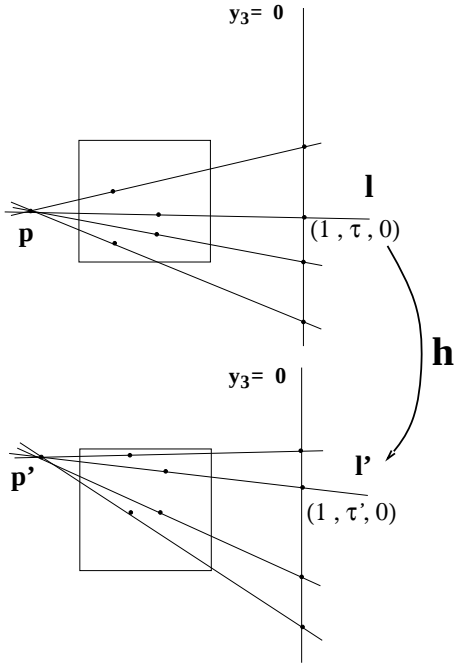

Figure 3: Parameterization of the epipolar transformation

Note that in the case where one of the epipole is at infinity, the previous parameterization is no longer valid, but we can generalize it by considering instead of the lines at infinity, the general lines. A more complicated projective parameter τ (depending also on the choice of the lines $\langle p_1, p_2 \rangle$ and $\langle p'_1, p'_2 \rangle$) is obtained in the same way by intersecting an epipolar line with $\langle p_1, p_2 \rangle$ in the first image, $\langle p'_1, p'_2 \rangle$ in the second image, and then the relation (6) holds as well. The expression (7) becomes:

$$\begin{aligned} F_{ij} = &(\mathbf{p}'_1 \times \mathbf{e}')_i (\mathbf{p}_2 \times \mathbf{e})_j \, a - (\mathbf{p}'_1 \times \mathbf{e}')_i (\mathbf{p}_1 \times \mathbf{e})_j \, b \\ &+ (\mathbf{p}'_2 \times \mathbf{e}')_i (\mathbf{p}_2 \times \mathbf{e})_j \, c - (\mathbf{p}'_2 \times \mathbf{e}')_i (\mathbf{p}_1 \times \mathbf{e})_j \, d \end{aligned}$$

2.5 Euclidean interpretation: Relation with Longuet-Higgins equation

The Longuet-Higgins equation [40] applies when using normalized coordinates, and thus calibrated cameras. In that case, the 2-D projective coordinates of a pixel $\mathbf{m}$ are equivalent to the 3-D direction of the optical ray $\langle C, m \rangle$, which is of course not the case with retinal (uncalibrated) coordinates. If the motion between the two positions of the cameras is given by the rotation matrix $\mathbf{R}$ and the translation vector $\mathbf{t}$, and if m and m' are corresponding points, then the fact that the three vectors $\mathbf{Cm}'$, $\mathbf{t}$, and $\mathbf{Cm}$ are coplanar can be written as:

$$\mathbf{m}' \cdot (\mathbf{t} \times \mathbf{Rm}) = \mathbf{m}'^{T} [\mathbf{t}]_{\times} \mathbf{Rm} = \mathbf{m}'^{T} \mathbf{Em} = 0 \tag{9}$$

We have replaced the vectors $\mathbf{Cm}'$ and $\mathbf{Cm}$ with $\mathbf{m}'$ and $\mathbf{m}$, respectively since, according to a remark of the previous section, they are proportional.

The matrix $\mathbf{E}$, which is the product of an orthogonal matrix and an antisymmetric matrix, is called an Essential matrix. Because of the depth/speed ambiguity, $\mathbf{E}$ depends on five parameters

only, the three parameters of the 3-D rotation, and the two parameters defining the direction of translation.

It can be seen that the equation (9) is a special case of (2). Since normalized coordinates (used in (9)) are obtained from pixel coordinates (used in (2)) by a multiplication by the inverse of the intrinsic parameters matrix $\mathbf{A}$, we have the relation:

$$\mathbf{F} = \mathbf{A}^{-1T}\mathbf{E}\mathbf{A}^{-1} \tag{10}$$

Unlike the essential matrix, which is characterized by the two constraints found by Huang and Faugeras [30] which are the nullity of the determinant and the equality of the two non-zero singular values, the only property of the Fundamental matrix is that it is of rank two. As it is also defined only up to a scale factor, the number of independent coefficients of $\mathbf{F}$ is seven. We will see in section 4.4 that the Fundamental matrix can be written as a product of an invertible matrix and an antisymmetric matrix.

2.6 Summary

In this section, we have described the epipolar transformation, both from a geometrical point of view and from an algebraic point of view. In order to provide the latter one, we have defined the *Fundamental matrix*. Its properties and relations to the well-known *Essential matrix* have been made clear. It must be noted that the Fundamental matrix provides a complete description of the projective structure of a set of two cameras. No other geometrical information can be obtained from uncalibrated cameras without making further assumptions about the structure of the scene.

3 General estimation methods: an analysis and experimental results

3.1 The linear criterion

The eight point algorithm Equation 2) is linear and homogeneous in the 9 unknown coefficients of matrix $\mathbf{F}$. Thus we know that if we are given 8 matches we will be able, in general, to determine a unique solution for $\mathbf{F}$, defined up to a scale factor. This approach, known as the eight point algorithm, was introduced by Longuet-Higgins [40] and has been extensively studied in the literature [38] [75] [12] [79] [37], for the computation of the *Essential matrix*. It has proven to be very sensitive to noise. Our contribution is to study it in the more general framework of *Fundamental matrix* computation. Some recent work has indeed pointed out that it is also relevant for the purpose of working from uncalibrated cameras [57], [16] [23]. In this framework, we obtain new results about the accuracy of this criterion, which will enable us to present a more robust approach.

In practice, we are given much more than 8 matches and we use a least-squares method to solve:

$$\min_{\mathbf{F}} \sum_i (\mathbf{q}_i'^T \mathbf{F} \mathbf{q}_i)^2 \tag{11}$$

When a Fundamental matrix obtained numerically does not verify the rank constraint, there is no exact solution to $\mathbf{F}\mathbf{e} = 0$. In that case, we cannot use formulas (7), thus the epipole e is determined by solving the following classical constrained minimization problem

$$\min_{\mathbf{e}} \|\mathbf{F}\mathbf{e}\|^2 \quad \text{subject to} \quad \|\mathbf{e}\|^2 = 1 \tag{12}$$

which yields $\mathbf{e}$ as the unit norm eigenvector of matrix $\mathbf{F}^T\mathbf{F}$ corresponding to the smallest eigenvalue. The same processing applies in reverse to the computation of the epipole $\mathbf{e}'$. The epipolar transformation can then be obtained by a linear least-squares procedure, using equations (5) and (6).

The advantage of the linear criterion is that it leads to a non-iterative computation method, however, we have found that it is quite sensitive to noise, even with numerous data points. Let us point out to the two main drawbacks of the linear criterion. A more detailed analysis of the linear criterion is performed in [43, 42], where some analytical results and numerical examples are provided.

The linear criterion cannot express the rank constraint Let l' be an epipolar line in the second image, computed from a fundamental matrix $\mathbf{F}$ that was obtained by the linear criterion, and from the point $\mathbf{m} = (u, v, 1)^T$ of the first image. We can express $\mathbf{m}$ using the epipole in the first image, and the horizontal and vertical distances from this epipole, x and y. A projective representation for l' is:

$$\mathbf{l}' = \mathbf{F}\mathbf{m} = \mathbf{F}\begin{pmatrix} e_1 - x \\ e_2 - y \\ 1 \end{pmatrix} = \mathbf{F}\mathbf{e} - \underbrace{\mathbf{F}\begin{pmatrix} x \\ y \\ 0 \end{pmatrix}}_{\mathbf{l}_1} \tag{13}$$

If $\det(\mathbf{F}) = 0$, the epipole e satisfies exactly $\mathbf{F}\mathbf{e} = 0$, thus the last expression simplifies to $\mathbf{l}_1$, which is an epipolar line. If the determinant is not exactly zero, we see that $\mathbf{l}'$ is the sum of a constant vector $\mathbf{r} = \mathbf{F}\mathbf{e}$ which should be zero but is not, and of the vector $\mathbf{l}_1$, whose norm is bounded by $\sqrt{x^2 + y^2}\|\mathbf{F}\|$. We can conclude that when $(x, y) \to (0, 0)$ $(m \to e)$, the epipolar line of m in the second image converges towards a fixed line represented by $\mathbf{r}$, which is inconsistent with the notion of epipolar geometry. We can also see that the smaller $\sqrt{x^2 + y^2}$ is (ie the closer m is to the epipole), the bigger will be the error on its associated epipolar line. In particular, it can be concluded that *if the epipole is in the image, the epipolar geometry described by the fundamental matrix obtained from the linear criterion will be inaccurate.* This problem can be observed in the images shown in the experimental part, in figure 11 for the intersection of epipolar lines, and in figure 12, for the inconsistency of epipolar geometry near the epipoles.

The linear criterion suffers from lack of normalization Let us now give a geometrical interpretation of the criterion (11). The Euclidean distance of the point q' of the second image to the epipolar line $\mathbf{l}' = (l'_1, l'_2, l'_3)^T = \mathbf{F}\mathbf{q}$ of the corresponding point q of the first image is:

$$d(\mathbf{q}', \mathbf{l}') = \frac{|\mathbf{q}'^T \mathbf{l}'|}{\sqrt{(l'_1)^2 + (l'_2)^2}} \tag{14}$$

We note that this expression is always valid as the normalizing term $k = \sqrt{(l'_1)^2 + (l'_2)^2}$ is null only in the degenerate cases where the epipolar line is at infinity. The criterion (11) can be written:

$$\sum_i k_i^2 d^2(\mathbf{q}'_i, \mathbf{l}'_i) \tag{15}$$

This interpretation shows that a geometrically significant quantity in the linear criterion is the distance of a point to the epipolar line of its corresponding point. This quantity is weighted by the coefficients k, defined above. To see why it can introduce a bias, let us consider the case where the displacement is a pure translation. The fundamental matrix is antisymmetric and has the form:

$$\begin{bmatrix} 0 & 1 & -y \\ -1 & 0 & x \\ y & -x & 0 \end{bmatrix}$$

where $(x, y, 1)^T$ are the coordinates of the epipoles, which are the same in the two images. If $(u_i, v_i, 1)^T$ are the coordinates of the point q_i in the first image, then the normalizing factor is $k_i^2 = \lambda^2((y - v_i)^2 + (x - u_i)^2)$, where λ is a constant. When minimizing the criterion (15), we will minimize both k_i and $d^2(\mathbf{q}'_i, \mathbf{l}'_i)$. But minimizing k_i is the same than favoring the fundamental

matrices which yield epipoles near the image. This is in fact still valid in the general case [42], and we conclude that *the linear criterion shifts epipoles towards the image center.*

3.2 Non-Linear criteria

The distance to epipolar lines We now introduce a first non-linear approach, based on the geometric interpretation of criterion (11) given in 3.1. To obtain a consistent epipolar geometry, it is necessary and sufficient that by exchanging the two images, the fundamental matrix is changed to its transpose. This yields the following symmetric criterion:

$$\sum_i (d^2(\mathbf{q}'_i, \mathbf{F}\mathbf{q}_i) + d^2(\mathbf{q}_i, \mathbf{F}^T\mathbf{q}'_i))$$

and can be written, using (14) and the fact that $\mathbf{q}_i^{'T}\mathbf{F}\mathbf{q}_i = \mathbf{q}_i^T\mathbf{F}^T\mathbf{q}'_i$:

$$\sum_i \left(\frac{1}{(\mathbf{F}\mathbf{q}_i)_1^2 + (\mathbf{F}\mathbf{q}_i)_2^2} + \frac{1}{(\mathbf{F}^T\mathbf{q}'_i)_1^2 + (\mathbf{F}^T\mathbf{q}'_i)_2^2} \right) (\mathbf{q}_i^{'T}\mathbf{F}\mathbf{q}_i)^2 \tag{16}$$

This criterion, which will be referred to in the sequel by **DIST** is clearly normalized in the sense that it does not depend on the scale factor used to compute $\mathbf{F}$.

The Gradient criterion and an interpretation as a distance Correspondences are obtained with some uncertainty. When minimizing the expression (11), we have a sum of terms $\mathcal{E}_i = \mathbf{q}_i^{'T}\mathbf{F}\mathbf{q}_i$ which have different variances. It is natural to weight them so that the contribution of each of these terms to the total criterion will be inversely proportional to its variance. The variance of $\mathcal{E}_i$ is given as a function of the variance of the points q_i et q'_i by:

$$\sigma^2_{\mathcal{E}_i} = \begin{bmatrix} \frac{\partial \mathcal{E}_i^T}{\partial \mathbf{q}_i} & \frac{\partial \mathcal{E}_i^T}{\partial \mathbf{q}'_i} \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda}_{\mathbf{q}_i} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_{\mathbf{q}'_i} \end{bmatrix} \begin{bmatrix} \frac{\partial \mathcal{E}_i}{\partial \mathbf{q}_i} \\ \frac{\partial \mathcal{E}_i}{\partial \mathbf{q}'_i} \end{bmatrix} \tag{17}$$

where $\Lambda_{\mathbf{q}_i}$ and $\Lambda_{\mathbf{q}'_i}$ are the covariance matrices of the points q et q', respectively. These points are uncorrelated as they are measured in different images. We make the classical assumption that their covariance is isotropic and uniform, that is:

$$\Lambda_{\mathbf{q}_i} = \Lambda_{\mathbf{q}'_i} = \begin{bmatrix} \sigma & 0 \\ 0 & \sigma \end{bmatrix}$$

The equation (17) reduces to:

$$\sigma^2_{\mathcal{E}_i} = \sigma^2 \|\nabla \mathcal{E}_i\|^2$$

where $\nabla \mathcal{E}_i$ denotes the gradient of $\mathcal{E}_i$ with respect to the four-dimensional vector $(u_i, v_i, u'_i, v'_i)^T$ built from the affine coordinates of the points q_i and q'_i. Thus:

$$\nabla \mathcal{E}_i = ((\mathbf{F}^T\mathbf{q}'_i)_1, (\mathbf{F}^T\mathbf{q}'_i)_2, (\mathbf{F}\mathbf{q}_i)_1, (\mathbf{F}\mathbf{q}_i)_2)^T$$

We obtain the following criterion, referred to in the sequel as **GRAD**, which is also normalized:

$$\sum_i \frac{(\mathbf{q}_i^{'T}\mathbf{F}\mathbf{q}_i)^2}{(\mathbf{F}\mathbf{q}_i)_1^2 + (\mathbf{F}\mathbf{q}_i)_2^2 + (\mathbf{F}^T\mathbf{q}'_i)_1^2 + (\mathbf{F}^T\mathbf{q}'_i)_2^2} \tag{18}$$

We can note that there is a great similarity between this criterion and the distance criterion (16). Each of its terms has the form $\frac{1}{k^2+k'^2}\mathcal{E}$, whereas the first one has terms $(\frac{1}{k^2} + \frac{1}{k'^2})\mathcal{E}$.

We can also consider the problem of the computing the fundamental matrix from the definition (2) in the general framework of surface fitting, The surface $\mathcal{S}$ is modeled by the implicit equation $g(\mathbf{x}, \mathbf{f}) = 0$, where $\mathbf{f}$ is the sought parameter vector describing the surface which best

fits the data points $\mathbf{x}_i$. The goal is to minimize a quantity $\sum_i d(\mathbf{x}_i, \mathcal{S})^2$, where d is a distance. In our case, the data points are the vectors $\mathbf{x}_i = (u_i, v_i, u'_i, v'_i)$, $\mathbf{f}$ is one of the 7 dimensional parameterizations introduced in the previous section,, and g is given by (2). We have developed a method to perform the *exact* computation of this distance [42, 43], based on some special properties of the surface $\mathcal{S}$, but this approach is computationally very expensive.

The linear criterion can be considered as a generalization of the Bookstein distance [5] for conic fitting. The straightforward idea is to approximate the true distance of the point $\mathbf{x}$ to the surface by the number $g(\mathbf{x}, \mathbf{f})$, in order to get a closed-form solution. A more precise approximation has been introduced by Sampson [64]. It is based on the first-order approximation:

$$g(\mathbf{x}) \simeq g(\mathbf{x}_0) + (\mathbf{x} - \mathbf{x}_0) \cdot \nabla g(\mathbf{x}) = g(\mathbf{x}_0) + \|\mathbf{x} - \mathbf{x}_0\| \, \|\nabla g(\mathbf{x})\| \cos(\mathbf{x} - \mathbf{x}_0, \nabla g(\mathbf{x}))$$

If $\mathbf{x}_0$ is the point of $\mathcal{S}$ which is the nearest from $\mathbf{x}$, we have the two properties $g(\mathbf{x}_0) = 0$ and $\cos(\mathbf{x} - \mathbf{x}_0, \nabla g(\mathbf{x}_0)) = 1$. If we make the further first-order approximation that the gradient has the same direction at $\mathbf{x}$ and at $\mathbf{x}_0$: $\cos(\mathbf{x} - \mathbf{x}_0, \nabla g(\mathbf{x}_0)) \simeq \cos(\mathbf{x} - \mathbf{x}_0, \nabla g(\mathbf{x}))$, we get:

$$d(\mathbf{x}, \mathcal{S}) = \|\mathbf{x} - \mathbf{x}_0\| \simeq \frac{g(\mathbf{x})}{\|\nabla g(\mathbf{x})\|}$$

Il is now obvious that the criterion (18) can be written: $\sum_i d(\mathbf{x}_i, \mathcal{S})^2$.

3.3 Parameterizations of the Fundamental Matrix

A matrix defined up to a scale factor The most natural idea to take into account the fact that $\mathbf{F}$ is defined only up to a scale factor is to fix one of the coefficients to 1 (only the linear criterion allows us to use in a simple manner another normalization, namely $\|\mathbf{F}\|$). It yields a parameterization of $\mathbf{F}$ by eight values, which are the ratio of the eight other coefficients to the normalizing one.

In practice, the choice of the normalizing coefficient has significant numerical consequences. As we can see from the expressions of the criteria previously introduced (16) and (18), the non-linear criteria take the general form:

$$\frac{Q_1(F_{11}, F_{12}, F_{13}, F_{21}, F_{22}, F_{23}, F_{31}, F_{32}, F_{33})}{Q_2(F_{11}, F_{12}, F_{13}, F_{21}, F_{22}, F_{23})}$$

where Q_1 and Q_2 are quadratic forms which have null values at the origin. A well-known consequence is that the function Q_1/Q_2 is not regular near the origin. As the derivatives are used in the course of the minimization procedure, this will induce unstability. As a consequence, we have to choose as normalizing coefficients one of the six first one, as only these coefficients appear in the expression of Q_2. Fixing the value of one of these coefficients to one prevents Q_2 from getting near the origin.

We have established using covariance analysis that the choices are not equivalent when the order of magnitude of the different coefficients of $\mathbf{F}$ is different. The best results are theoretically obtained when normalizing with the biggest coefficients. We found in our experiments this observation to be generally true. However, as some cases of divergence during the minimization process sometimes appear, the best is to try several normalizations

We note that as the matrices which are used to initialize the non-linear search are not, in general, singular, we have to compute first the closest singular matrix, and then the parameterization.

A singular matrix As seen in part 3.1, the drawback of the previous method is that we do not take into account the fact that the rank of $\mathbf{F}$ is only two, and that $\mathbf{F}$ thus depends on only 7 parameters. We have first tried to use minimizations under the constraint $\det(\mathbf{F}) = 0$, which is a cubic polynomial in the coefficients of $\mathbf{F}$. The numerical implementations were not efficient and accurate at all.

Thanks to a suggestion by Luc Robert, we can express the same constraint with an unconstrained minimization: the idea is to write matrix $\mathbf{F}$ as:

$$\mathbf{F} = \begin{pmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7a_1 + a_8a_4 & a_7a_2 + a_8a_5 & a_7a_3 + a_8a_6 \end{pmatrix} \tag{19}$$

The fact that the third line is a linear (thus this parameterization will be designated in the sequel by the letter $\mathbf{L}$) combination of the two first lines ensures that $\mathbf{F}$ is singular. Choosing such a representation allows us to represent $\mathbf{F}$ by the right number of parameters, once the normalization is done. A non-linear procedure is required, but it is not a drawback, as the criteria presented in section 3.2 are already non-linear.

A Fundamental matrix with finite epipoles The previous representation takes into account only the fact that $\mathbf{F}$ is singular. We can use the fact it is a Fundamental matrix to parameterize it by the values that are of significance for us, those defining the epipolar transformation (thus this parameterization will be designated in the sequel by the letter $\mathbf{T}$). Using the formulas (7) yield:

$$\mathbf{F} = \begin{pmatrix} b & a & -ay - bx \\ -d & -c & cy + dx \\ dy' - bx' & cy' - ax' & -cyy' - dy'x + ayx' + bxx' \end{pmatrix} \tag{20}$$

The parameters that we use are the affine coordinates (x, y) and (x', y') of the two epipoles, and three of the four homography coefficients, which are the coefficients of the submatrix 2×2 obtained by suppressing the third line and the third column. We normalize by the biggest of them. The initial parameters are obtained by computing the epipoles and the epipolar transformation by the approximations introduced in 2.4

3.4 An experimental comparison

We have presented an approach to the computation of the fundamental matrix which involves several parameterizations and several criteria. The goal of this part is to provide a statistical comparison of the different combinations.

The method An important remark is that if we want to make a precise assessment of the performance of any method, we have to change not only the image noise, as it is often done, but also the displacements. Different displacements will give rise to configurations with stability properties that are very different.

We start from 3D points that are randomly scattered in a cube, and from a projection matrix $\mathbf{P}$. All these values are chosen to be realistic. Each trial consists of:

- Take a random rigid displacement $\mathbf{D}$,
- Compute the exact fundamental matrix $\mathbf{F}_0$ from $\mathbf{D}$ and $\mathbf{P}$,
- Compute the projection matrix $\mathbf{P}'$ from $\mathbf{D}$ and $\mathbf{P}$,
- Project the 3D points in the two 512×512 retinas using $\mathbf{P}$ and $\mathbf{P}'$,
- Add Gaussian noise to the image points,
- Solve for the fundamental matrix $\mathbf{F}$,
- Compute the relative distance of the epipoles from $\mathbf{F}$ and those from $\mathbf{F}_0$.

In many applications (see the last section of this paper), only the coordinates of the epipoles are needed. In some sense, they are the most important piece of information contained in the

Fundamental matrix, and thus it is natural to attempt to quantify errors on this matrix by errors on its epipoles. We define the relative error, for each coordinate of the epipole:

$$\min\{\frac{|x - x_0|}{\min(|x|, |x_0|)}, 1\}$$

We took a relative error since a same (absolute) deviation of the epipole will yield a larger error on the epipolar geometry if the epipole lies closer to the image center. This allows us to ensure a consistent maximal error on the direction of epipolar lines, regardless of the distance of the epipole to the image center. It has to be noted that the choice of an error measure for the epipoles is not an obvious matter, since they are basically quantities of the projective plane, which has no metric. A further discussion of error measures can be found in [46]. All the graphs shown in this section are averaged over a few hundred trials. In a scheme where only such a small number of experiments are averaged, a single very large value could significantly affect statistics, and this is why the relative error is thresholded by 1.

As our experimentations have shown that the average errors on the four coordinates are always coherent, we will take the mean of these four values as an error measure. Some experimental evidence to show that it is indeed an adequate characterization is provided next.

Epipoles stability characterize Fundamental matrix stability The estimation of the fundamental matrix can be done as a two-stage process, the first one being the estimation of the coordinates of the epipoles, and the second one the estimation of the coefficients of the homography. If one of the two stages is significantly more sensitive to noise than the other one, then we can conclude that its stability determines the stability of the overall estimation.

- The fundamental matrix has been computed from point correspondences using the quadratic criterion derived from the linear relation (2). The epipoles e and e' are then computed from this matrix using (12).
- The coefficients of the epipolar homography have been computed from the point correspondences and the *correct* epipoles, using a linear least-squares formulation based on the relation derived by making substitutions of (5) in (6).

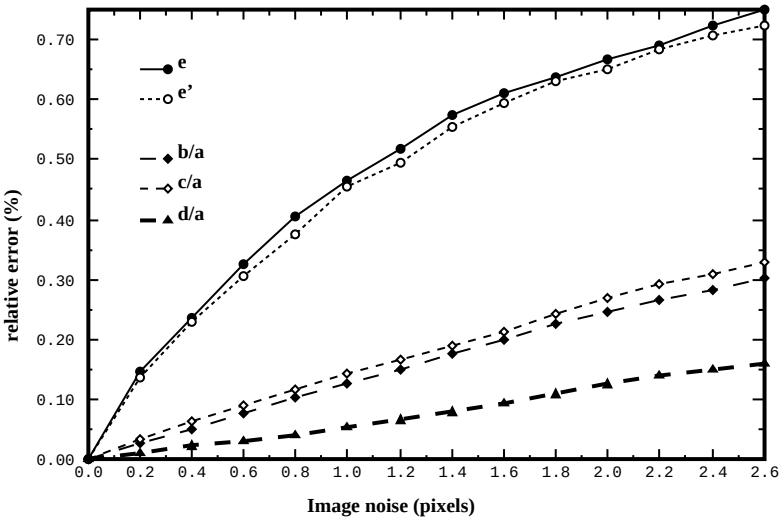

Figure 4: Sensitivity to noise of the different components of the fundamental matrix.

Since the four coefficients of the epipolar transformation are defined only up to a scale factor, we have normalized them by dividing by a, which allows to consider a relative error for each of them. From the results of the simulation shown Fig. 4, it is clear that:

- The stability of the epipoles in each of the images is comparable, which was to be expected, since the criterion (2) is symmetrical. Note that the non-linear criteria proposed in [43] also share this property.
- Once the epipoles are determined correctly, the computation of the homography is quite stable, and thus that the more unstable part of the computation is the determination of the epipoles.

We thus conclude from this simulation that *an adequate measure for the stability of the fundamental matrix is the stability of one of its epipoles.* Note that this is consistent with the findings of [47], where it has been shown that the epipole plays a particular role in the projective description of the geometry of a system of two cameras.

Non-linear criteria There are two different parameterizations, that are presented Section 3.3, and two different non-linear criteria, presented in Section 3.2. The abbreviations for the four resulting combinations that we studied are in table 1. We have tried several minimization procedures, including material from Numerical Recipes, and programs from the NAG library.

Table 1: Non-linear methods for the computation of the fundamental matrix

abbrev.	criterion	parameterization
LIN	linear	normalization by $\|\mathbf{F}\|$
DIST-L	distance to epipolar lines (16)	singular matrix (19)
DIST-T	distance to epipolar lines	epipolar transformation (20)
GRAD-L	weighting by the gradient (18)	singular matrix
GRAD-T	weighting by the gradient	epipolar transformation

The comparison we have done is threefold:

1. *The stability of the minimum corresponding to the exact solution.* When noise is present, the hypersurface which represents the value of the criterion as a function of the parameters gets distorted, thus the coordinates of the minimum change. A measure of this variation is given by the distance between the exact epipole and the one obtained when starting the minimization with the exact epipole (figure 5).
2. *The convergence properties.* The question is whether it is possible to obtain a correct result starting from a plausible initialization, the matrix obtained from the linear criterion. We thus measure the distance between the exact epipole and the one obtained when starting the minimization with the linear solution (figure 6), and the distance between the epipole obtained when starting the minimization with the exact epipole and the one obtained when starting the minimization with the linear solution (figure 7) .
3. *The stability of the criterion.* When the hypersurface which represents the value of the criterion as a function of the parameters gets distorted, the values of the criterion at local minima corresponding to inexact solutions can become less than the value of the criterion at the correct minimum (figure 8).

The conclusions are:

- *The non-linear criteria are always better than the linear criterion.* When starting a non-linear computation with the result of the linear computation, we always improve the precision of the result, even if the noise is not important. The difference increases with the noise.
- The difference due to the choice of the criterion (**DIST** or **GRAD**, defined in section3.2) is much less significant than the one due to the choice of the parameterization (**L** or **T**, defined in section 3.3).

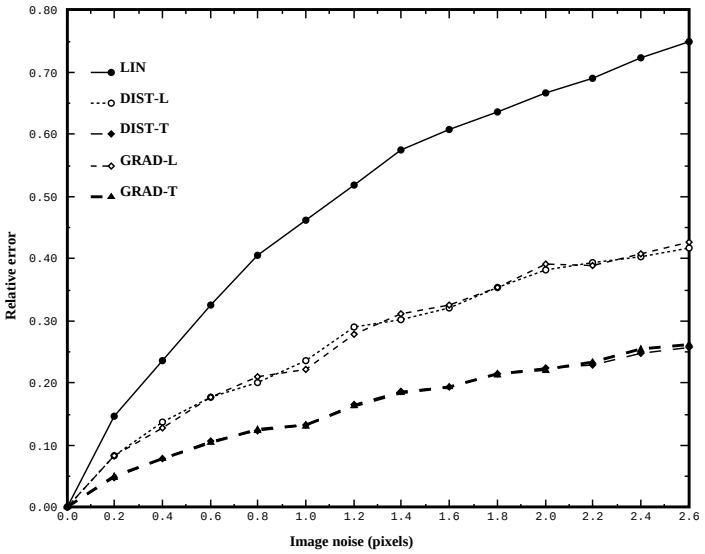

Figure 5: Relative errors obtained starting from the exact values.

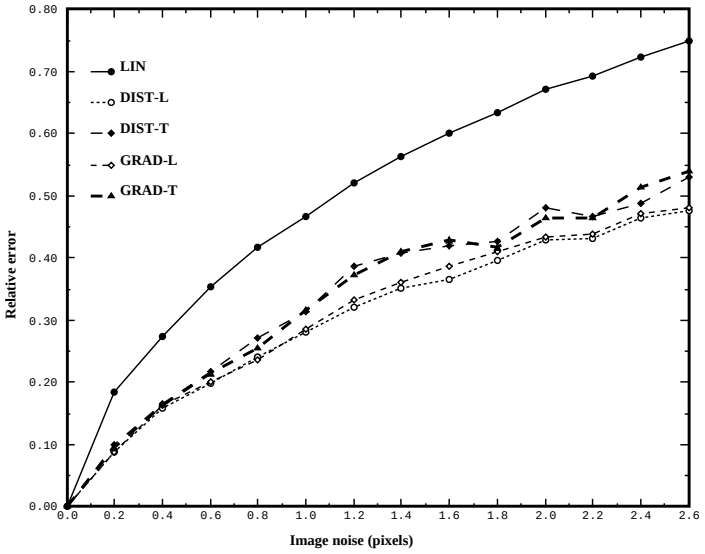

Figure 6: Relative errors obtained starting from the values found by the linear criterion.

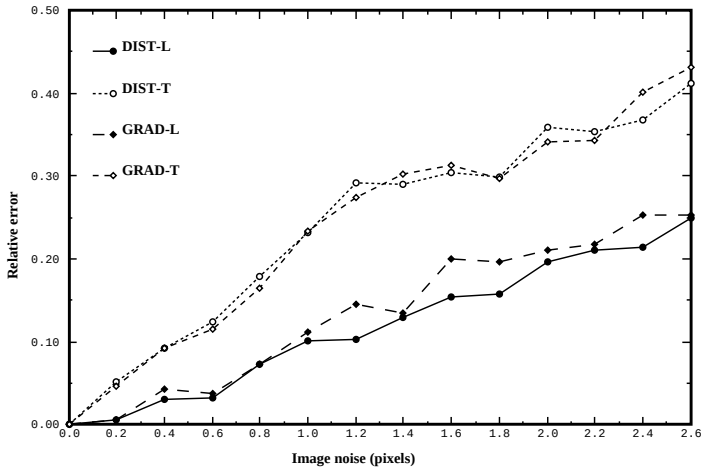

Figure 7: Relative distances between epipoles obtained by a minimization started from the exact value and epipoles obtained by a minimization started from the values found by the linear criterion.

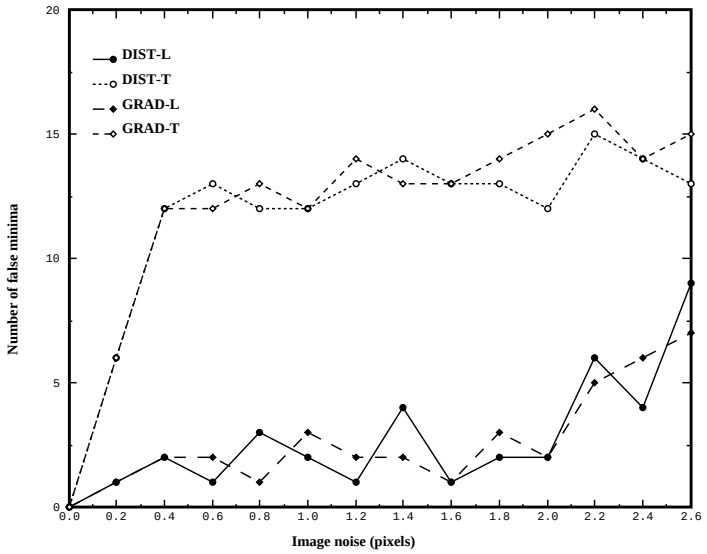

Figure 8: Number of false minima.

- The parameterization **T** yields more stable minima than the parameterization **L**, as seen in figure 5.
- However, the criterion obtained with parameterization **T** has worse convergence and stability properties than the parameterization **L** , as seen in figures 7 and 8
- As a consequence, when starting from the results of the linear criterion, the results of the four non-linear combinations are roughly equivalent, the results obtained with the parameterization **L** and the criterion **DIST** being slightly better, as seen in figure 6.
- The computation is quite sensitive to pixel noise: a Gaussian noise of variance 1 pixel yields a relative error which is about 30%.

Real data We now illustrate the remarks made in section 3.1 with a pair of images. It can be seen in figure 9 that the pencils of epipolar lines obtained with the linear criterion, and those obtained with the non-linear criterion are very different. The epipoles obtained with the non-linear criterion are much further away. It seems at first that if one considers a point that was used in the computation, its epipolar line lies very close to its corresponding point. However, the zoom of figure 10 shows that the fit is significantly better with the non-linear criterion. Figure 11 shows a set of epipolar lines obtained from the linear criterion, we can see that they don't meet exactly at a point, whereas they do by construction for the non-linear criterion. A consequence is illustrated in figure 12, which shows some more epipolar lines, drawn from points that were *not* used in the computation of the fundamental matrix. It can be seen that for the points on the wall, which are quite far from the epipole, the corresponding epipolar lines seem approximately correct, while for the points chosen on the table, the corresponding epipolar lines are obviously very incorrect, in the sense they are very far from the corresponding points. This situation does not occur with the non-linear criterion, as it can be seen in the bottom of this figure.

3.5 Summary

In this section, we focused on the problem of determining in a robust way the *Fundamental matrix* from a given number of image point correspondences. The classical linear criterion has been shown to be unable to express the rank and normalization constraints. Analyzing these drawbacks enabled us to introduce non-linear computation techniques, based on criteria that have a nice interpretation in terms of distances, and appropriate parameterizations. We have shown, using both large sets of simulations and real data, that our non-linear computation techniques provide significant improvement in the accuracy of the Fundamental matrix determination, and we have evaluated stability and convergence properties of each method.

4 Planes and the fundamental matrix: unstability and new algorithms

4.1 The correspondence between the two images of a plane

Definition Let M_i be space points which happen to lie in the same plane Π and m_i be their images by a projective linear relation from $\mathcal{P}^3$ to $\mathcal{P}^2$. Its restriction to Π is a projective linear relation between points of $\mathcal{P}^2$, which is an homography h. This relation is invertible, in the generic case. If two images of the points M_i lying in a plane, m_i and m'_i are available, we can consider the relation $h' \circ h^{-1}$ between these two images. It is thus an homography, which means there is a 3×3 invertible matrix $\mathbf{H}$, such that the following projective relation holds for each i:

$$\mathbf{m}'_i = \mathbf{H}\mathbf{m}_i \tag{21}$$

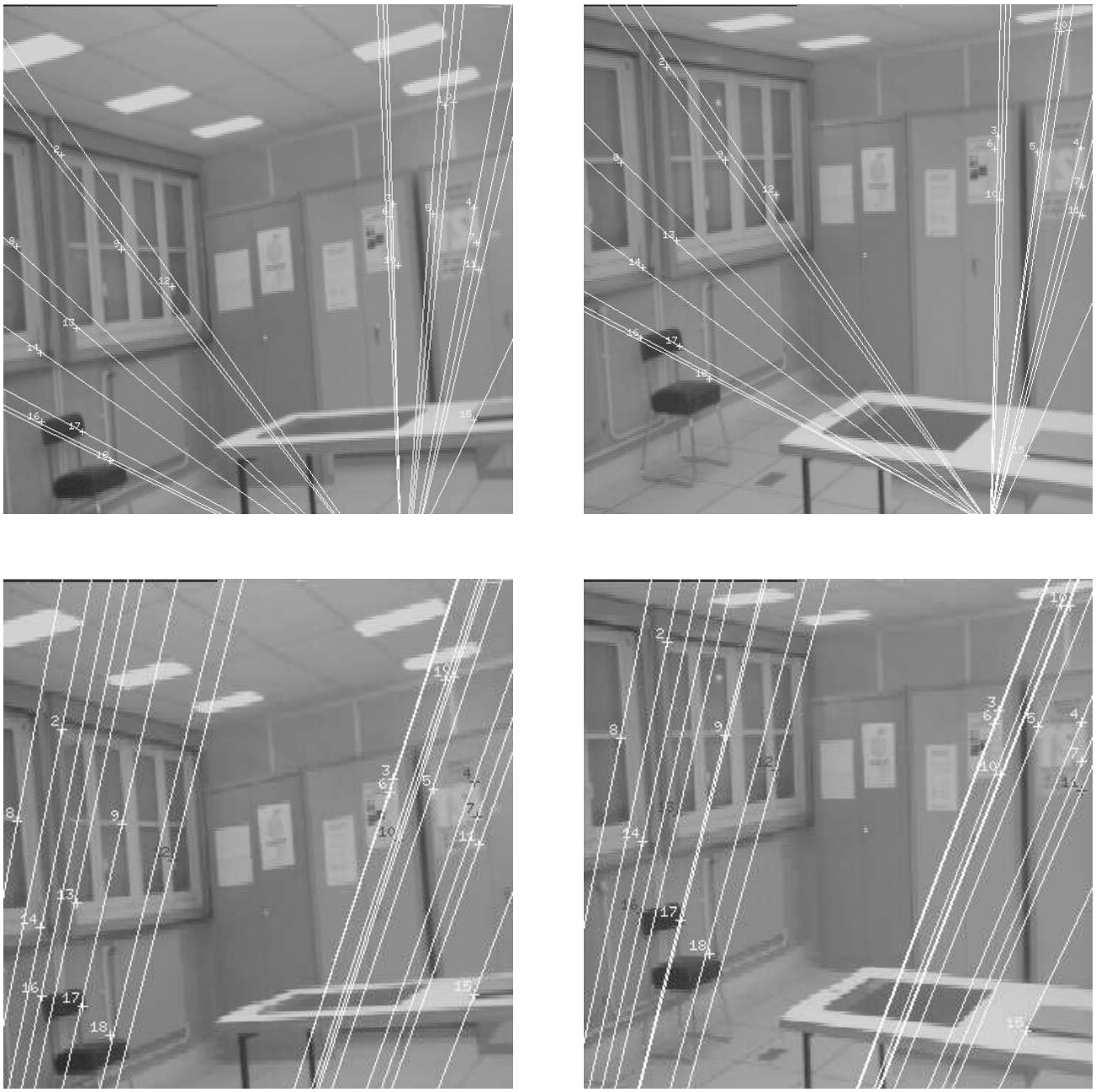

Figure 9: Epipolar lines obtained from the linear criterion (top), and from the non-linear criterion (bottom).

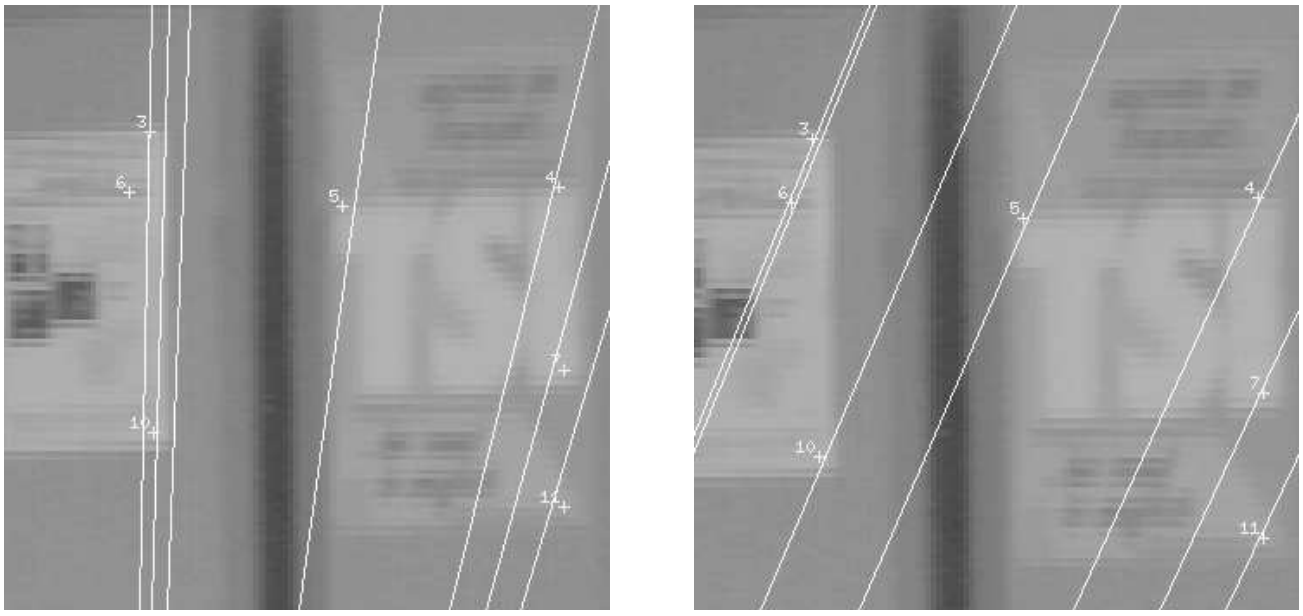

Figure 10: Zoom showing the fit with the linear criterion (left) and the non-linear criterion (right).

The fact that there is such an analytic relation between the coordinates of matched points entails that we are able to identify planes using only measurements in the image. Predict-and-verify algorithms have been already developed by [17], and more recently by [70] and [62], using uncalibrated cameras. The idea is to chose four points, to compute the homography, whose knowledge allows the position of the corresponding point of any new point on the plane to be predicted. The predicted position and the actual position are compared using a simple distance threshold, to decide whether the new point is on the plane defined by the four points. In this paper, we will not elaborate on this issue, but rather on the computational problems which can be solved once the identification of planes has been performed.

Computation We now study the problem of computing the parameters of the homography from point matches. By writing that the two proportionality constraints obtained from (21) are satisfied, we have two equations which are linear in the coefficients of $\mathbf{H}$, and can be solved as soon as four point correspondences are available. In practice about 10 to 20 points are needed to compute an accurate homography, and the (variable) number of points used in the experiment conducted in this section will never be less than that. We have found that with the criteria (**LIN**) based on a least-squares formulation of (21), there is a normalization problem, as for the computation of the fundamental matrix. A favorable thing is that we do not have a rank constraint to consider. The two non-linear criteria that we have investigated are similar to the ones introduced in Section 3.2:

DIST: symmeterized Euclidean distance between predicted and measured points,

GRAD: the two linear equations weighted by associated uncertainties.

Convergence properties have been tested by using two different initializations: the exact value (**-EX**) and the value obtained from the linear criterion (**-LIN**). In the statistical simulation shown figure 13, averaged over many widely different set of correspondences, the error measure is the average relative error on each coefficient. We can conclude from these results that:

- The non-linear criteria give better results than the linear criterion.

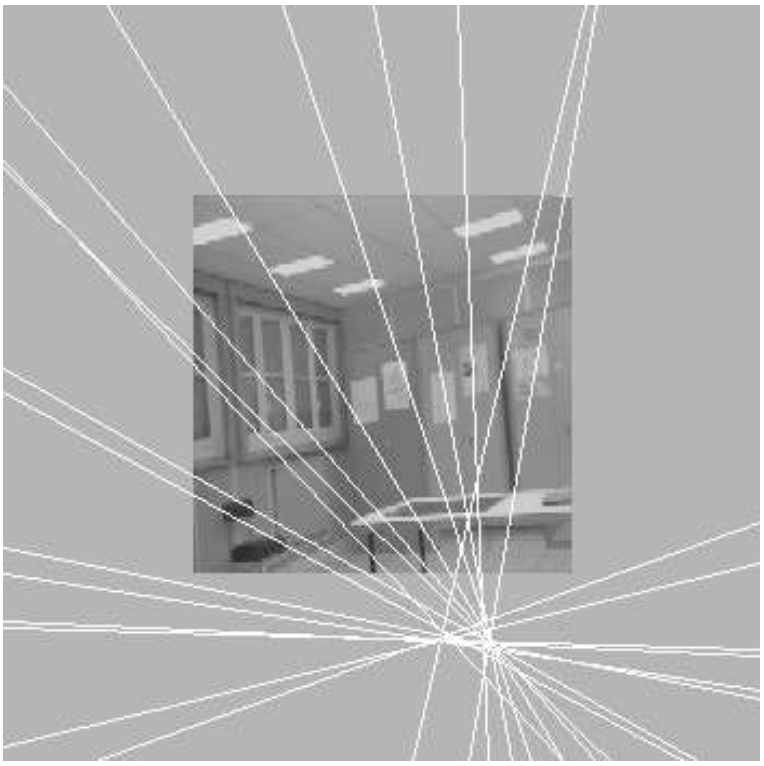

Figure 11: Intersection of epipolar lines obtained from the linear criterion.

- The results obtained from the two non-linear criteria are very close.
- The computation is more stable from the computation of the fundamental matrix, there is almost no convergence problem, thus it is possible to compute homographies with a good accuracy from point matches

4.2 Relation between homographies and the fundamental matrix

Let $\mathbf{F}$ be the fundamental matrix relating two images, and $\mathbf{H}$ an homography matrix relating the coordinates of points of a plane Π which projects in the two images. We consider the point m of the first image to be the projection of a virtual point M_Π of plane Π. The homography enables us to compute the projection m' of M_Π on the second image. The points m and m' are in correspondence, thus, using (2), we obtain:

$$\mathbf{m}'^T\mathbf{F}\mathbf{m} = (\mathbf{H}\mathbf{m})^T\mathbf{F}\mathbf{m} = \mathbf{m}^T\mathbf{H}^T\mathbf{F}\mathbf{m} = 0$$

This relation is to be satisfied by any point m, thus we can conclude that the matrix $\mathbf{H}^T\mathbf{F}$ is antisymmetric, which yields the important relation:

$$\mathbf{H}^T\mathbf{F} + \mathbf{F}^T\mathbf{H} = 0 \tag{22}$$

We are now going to show that a matrix $\mathbf{H}$ satisfies condition (22) if, and only if:

$$\mathbf{F} = [\mathbf{e}']_\times\mathbf{H} \tag{23}$$

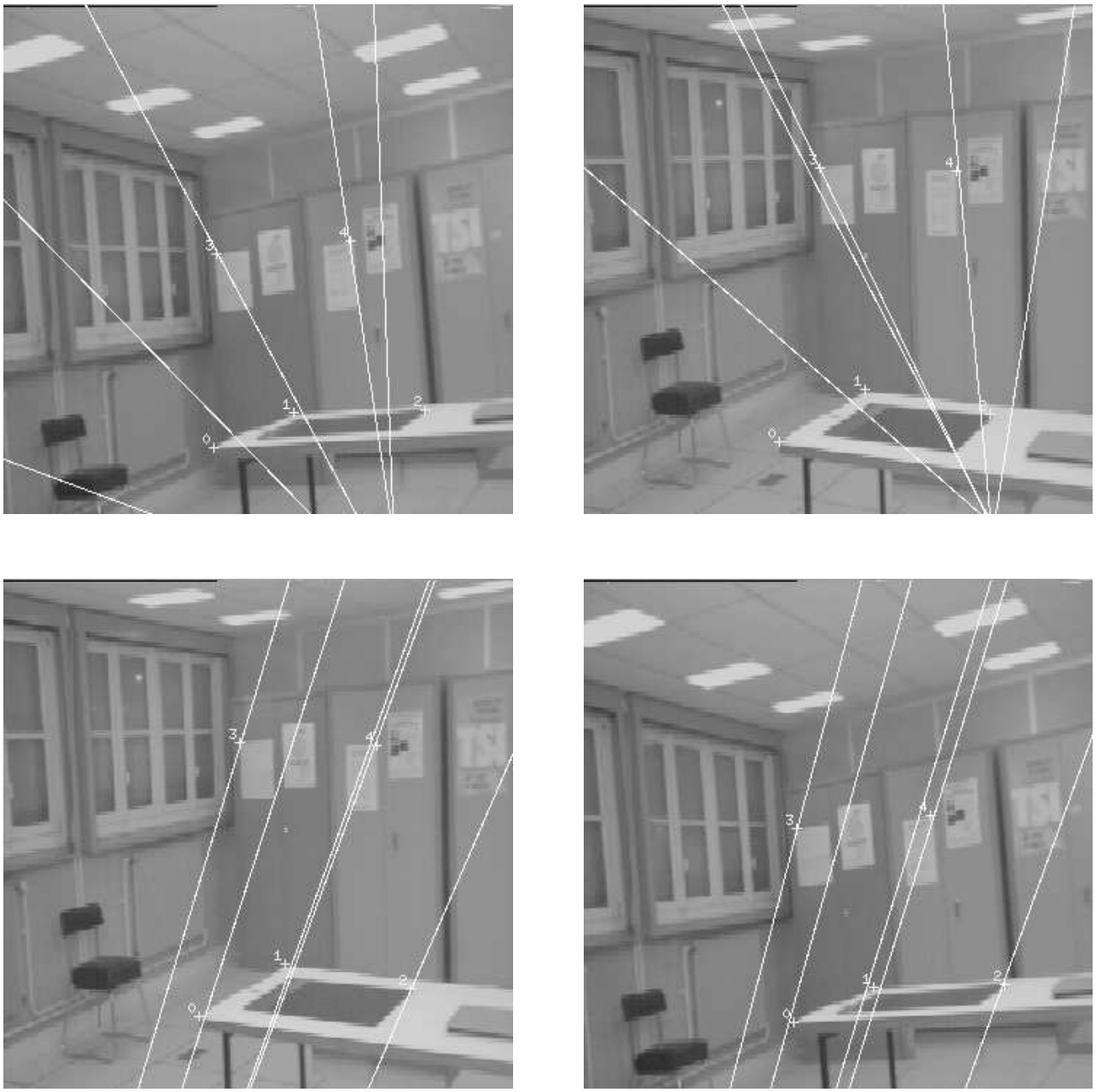

Figure 12: Additional epipolar lines obtained with the linear criterion (top), and with the non-linear criterion (bottom)

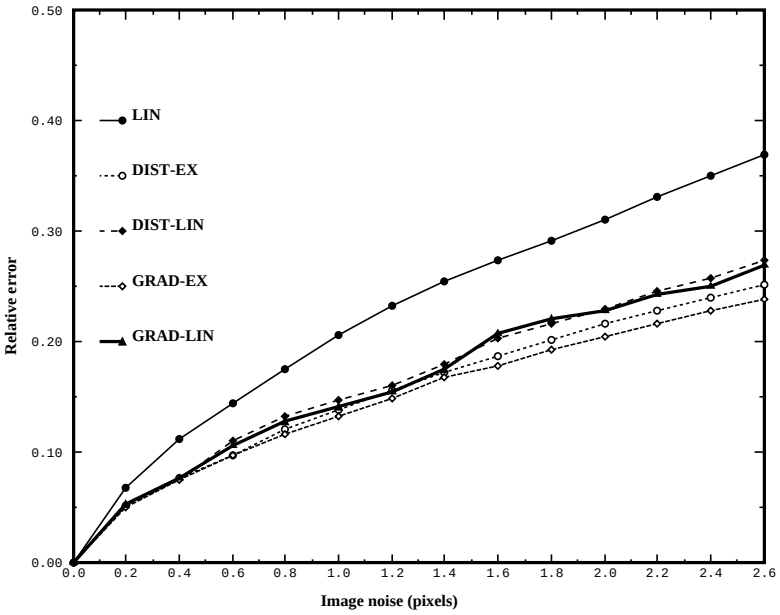

Figure 13: Relative errors on homography matrices computed with different methods.

It is straightforward to verify that (22) is satisfied if the substitution of (23) is made in that equation. Now let us suppose that $\mathbf{F}$ and $\mathbf{H}$ satisfy (22), and consider a pair of corresponding points m and m'. Since the fundamental matrix maps points to corresponding epipolar lines, $\mathbf{Fm} = \mathbf{e}' \times \mathbf{m}'$. Using this relation, we see that $(\mathbf{Fm})^T \mathbf{Hm} = 0$ (derived from (22)) is equivalent to: $\mathbf{m}'^T [\mathbf{e}']_\times \mathbf{Hm} = 0$. If we identify this equation with the epipolar constraint (2), we obtain the expression (23).

A first consequence of (22) is:

$$\mathbf{He} = \mathbf{e}' \quad \text{or} \quad \mathbf{He} = 0$$

The first case corresponds to the homography matrices, defined by a plane in general position. The second case corresponds to the degenerate case where the plane contains the optical center C', thus yielding a non-invertible correspondence. The value of the antisymmetric matrix $\mathbf{H}^T\mathbf{F}$ can also be obtained by remarking that the kernel of this matrix is the epipole e, thus:

$$\mathbf{H}^T\mathbf{F} \cong [\mathbf{e}']_\times$$

A third consequence of (22) is that the matrix $\mathbf{H}^T$ maps epipolar lines to corresponding epipolar lines, as also found by [27, 68]. Let l and l' be correspondent epipolar lines, containing respectively the points $\mathbf{m}$ and $\mathbf{m}' = \mathbf{Hm}$. We have, up to a scale factor:

$$\mathbf{H}^T\mathbf{l}' = \mathbf{H}^T\mathbf{Fm} = \mathbf{F}^T\mathbf{Hm} = \mathbf{F}^T\mathbf{m}' = \mathbf{l}$$

4.3 Planes and unstability of the fundamental matrix computations

The basic observation As a consequence of relation (22), we show in this section that if the points used for the computation of the fundamental matrix lie in a small number of planes, the computation will be unstable, even if it is theoretically well-defined as long as there are at least more than two planes.

We first consider the ideal case where the coordinates of the point matches are known exactly. If we have four point matches which come from the image of four points that lie in the same plane, a matrix $\mathbf{H}$ is determined. If we have an additional correspondence from the observation of a point that lies in the same plane, the coordinates $(\mathbf{m}_5, \mathbf{m}'_5)$ will verify $\mathbf{m}'_5 = \mathbf{Hm}_5$, and

since we have the relation (22), we will have $\mathbf{m}_5'\mathbf{F}\mathbf{m}_5 = 0$ *necessarily*. An interpretation of this fact is: whatever the number of correspondences we have, if they come from points which lie in the same plane, they will not produce more constraints on the matrix $\mathbf{F}$ than four points do. We see also that the constraints are actually the six equations which are derived from (22). Thus, in the case of a linear estimation of $\mathbf{F}$, where the rank constraint is ignored, there remains two undetermined parameters for $\mathbf{F}$, as shown previously by [18]. Taking into account this polynomial constraint would further restrain $\mathbf{F}$ to a one-parameter family.

This line of reasoning can be extended to the case of noisy data, by considering the covariance matrices [45, 42]

Experimental study: simulations We have first validated this result using an extensive simulation with noisy synthetic data. More detailed results are in [45, 42]. Each trial has been done as in section 3.4, except that at each time we have also chosen n random 3D points on p random planes. The conclusions are:

- When the number of planes is fixed, the quality of the estimation increases, as the number of points increases, which is classical.
- When the number of points is fixed, the quality of the estimation increases, as the number of planes in which they lie increases.
- The second variation is more important than the first one
- Better estimations are obtained with a few points that are on a several planes than with a great number of points which lie on a few planes. We see for instance that using 70 points on two planes is statistically worse than using only 10 points in general configuration.

Real cases The situation with two planes is not merely theoretical. We have encountered it several times, when trying to use our dihedral calibration grid to compute the fundamental matrix. In spite of the fact that corners can be detected with a subpixel accuracy on those grids, the localization of the epipoles were not accurate. An example of such images will be shown in the next section.

We now want to make a more general remark, which has very important practical consequences: even if the points are not exactly in an unstable configuration, if they are close enough to such configurations, we can observe ambiguity. Let take a real example: The upper part of figure 14 shows an image pair with corresponding points. Even though these points are not coplanar, we have estimated an homography from them using the method of section 4.1. The lower part of figure 14 shows the result of the mapping of the homography (respectively of its inverse) to the points of the first (respectively second) image. There is a very low difference between the real points and the points predicted by the homography. This fact indicates that while the 3D points are not coplanar, they can be approximated with a good precision by a plane. A consequence is that the epipolar geometry is ambiguous as expected, which can be seen on figure 15. The four frames represent the first image, with some epipolar lines. We can see that the position of the epipoles varies significantly from an image to another. Yet all the solutions give very low residuals (largely inferior to pixel) in terms of distance of corresponding point to epipolar lines. It is easy to test for an approximate single plane coplanarity using the residual of the computation of the homography. However, it is a little harder to test for an approximate bi-plane coplanarity, which is even more likely to happen. It will not yield ambiguity, as the two plane situation is not degenerate, but the precision of the estimation will suffer significantly.

4.4 Computation of the fundamental matrix from planes

A direct method to obtain the fundamental matrix The matrix equation (22) yields six scalar equations which are linear and homogeneous in the coefficients of $\mathbf{F}$. Let us suppose that we have determined two homographies, $\mathbf{H}_1$ et $\mathbf{H}_2$, from the observation of two

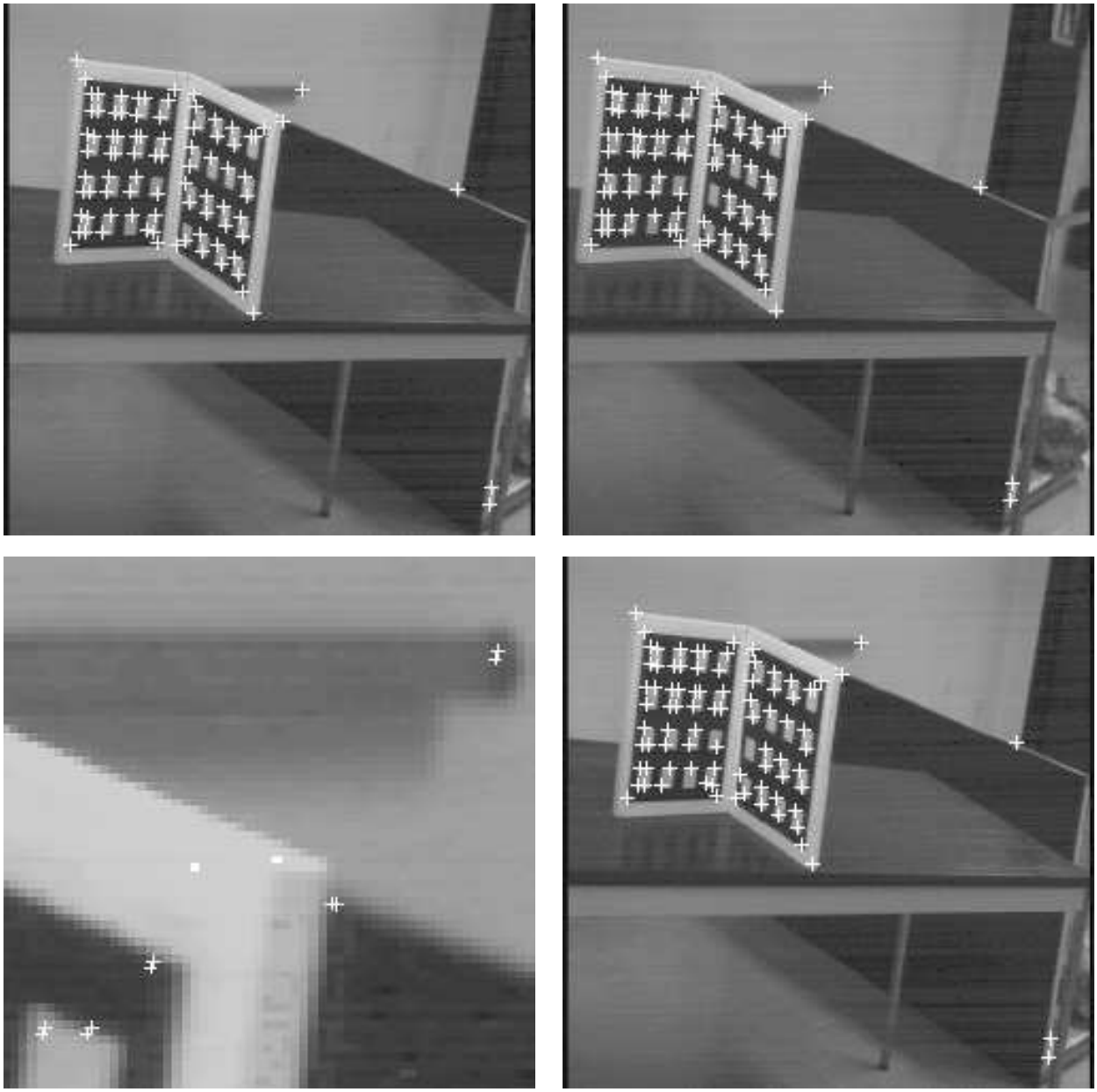

Figure 14: Actual matches (top), and matches predicted by the homography (bottom).

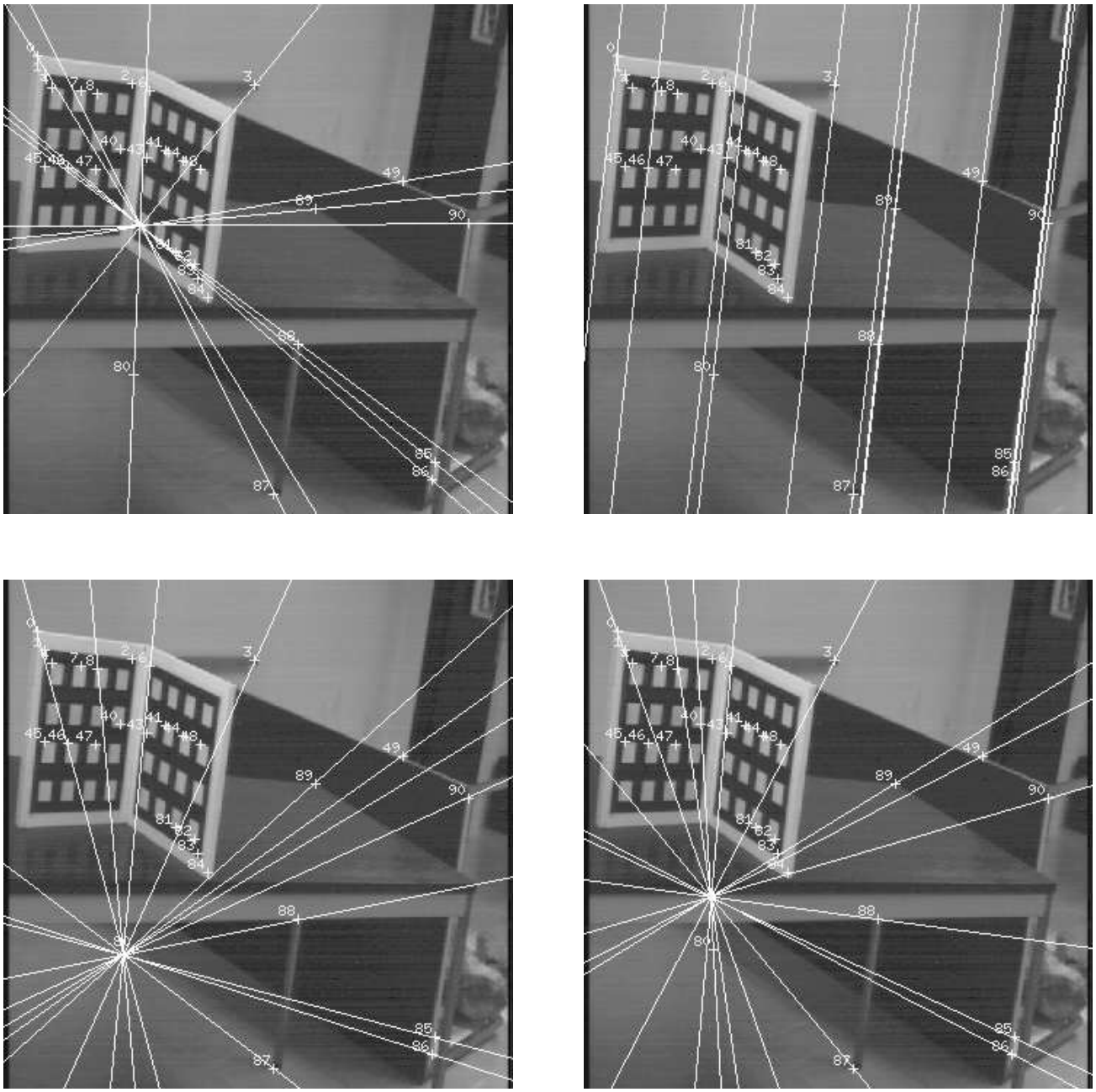

Figure 15: Ambiguity: four coherent epipolar pencils in the first image.

different plane areas. Putting together the two systems of equations (22), we obtain a system $\mathcal{S}$ of twelve equations which are linear and homogeneous in the coefficients of $\mathbf{F}$. Provided that we have identified at least two homographies, we can solve the $6N$ equations by a least-squares technique to obtain $\mathbf{F}$.

coordinates	e_x	e_y	e'_x	e'_y
EXACT	780.08	-427.04	1823.82	-1255.57
	LIN			
0.1 pixel	677.93	-867.29	2104.51	-1031.93
0.5 pixel	452.82	-1263.39	1628.43	59.84
1.0 pixel	330.67	-1567.84	1467.22	378.28
	NON-LIN-SYM			
0.1 pixel	782.70	-433.09	1845.02	-1283.58
0.5 pixel	781.70	-455.56	1870.00	-1362.63
1.0 pixel	760.32	-490.37	1887.33	-1436.37

Table 2: An example to illustrate the results obtained with different methods.

Two improvements can be made. The first is to into account the rank constraint on the fundamental matrix using one of the non-linear parameterizations introduced in section 3.3. The second one is to use a criterion, as explained in section 3.2, by minimizing simultaneously the equations (22) and the equations obtained by exchanging the role of the two images:

$$\mathbf{F}\mathbf{H}^{-1} + \mathbf{H}^{-1T}\mathbf{F}^T = 0$$

We also note that if we only have one plane, and the correspondence of at least two points which do not belong to the plane, we can obtain eight equations: six equations (22) and two equations (2), which is sufficient to compute $\mathbf{F}$.

A geometric interpretation: determining the epipoles We start from the following observation, which is illustrated in figure 16: let us suppose that we have two cameras, and that we observe in the second camera the images m'_1 and m'_2 of two space points M_1 and M_2, which have a same image by the first camera. Then $\langle m'_1, m'_2 \rangle$ is an epipolar line in the second image: as M_1 and M_2 have the same image by the first camera, the line $\langle M_1 M_2 \rangle$ is an optical ray for this camera. The line $\langle m'_1, m'_2 \rangle$, as the image of an optical ray of the first camera, is an epipolar line. This idea generalizes to the case of uncalibrated cameras with an arbitrary motion the notion of *motion parallax* [39]. In the original formulation of the idea, the two points M_1 and M_2 considered were physical points. This requires significant discontinuities in depth and identification of points which are aligned. By contrast, the idea behind all the methods that we will see next is that if we observe a plane, and if we know the homography relating the image of the points of the plane, we can *compute* such points m'_i from their images by the first camera m_i, by considering virtual points M_i. There are two possibilities to use this idea.

A first possibility is to use two planes Π_1 and Π_2. Let M be a space point, and m its image by the first camera. Let M_1 and M_2 be the points of Π_1 and Π_2, respectively, which image by the first camera is m. The image M_i in the second camera is $h_{\Pi_i}(m)$, with $i = 1, 2$. We thus obtain an epipolar line $\langle h_{\Pi_1}(m), h_{\Pi_2}(m) \rangle$ from each point m of the first image. For this approach it is necessary to identify two sets of four coplanar points, which define the homographies h_1 and h_2 (which matrices are respectively $\mathbf{H}_1$ and $\mathbf{H}_2$). The epipole e' is the point such that:

$$\forall \mathbf{m} \qquad \mathbf{e}'^T(\mathbf{H}_1\mathbf{m} \times \mathbf{H}_2\mathbf{m}) = 0$$

This equation defines a bilinear form of the coordinates of $\mathbf{m}$ and thus we can rewrite it as::

$$\forall \mathbf{m} \qquad \mathbf{m}^T\mathbf{B}\mathbf{m} = 0$$

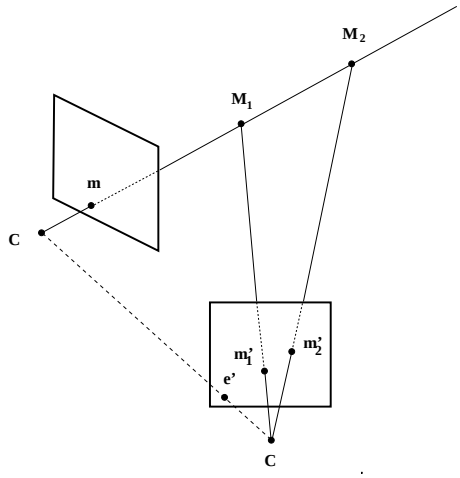

Figure 16: Epipolar line obtained by the observation of two special points.

where $\mathbf{B}$ is a 3×3 matrix depending quadratiquely of $\mathbf{H}_1$, $\mathbf{H}_2$, and linearly of $\mathbf{e}'$. From the last relation, we conclude that matrix $\mathbf{B}$ must be antisymmetric which gives a system $\mathcal{S}_0$ of six linear equations on the coordinates of the e', which we solve by a least squares method.

An interesting thing, that is not elaborated on here, is that these six equations should yield a rank two system, which yield four algebraic constraints on the homography matrices $\mathbf{H}_1$ and $\mathbf{H}_2$. We note that it was also the case with the system $\mathcal{S}$ where a system of twelve equations had to be of rank eight. Using the MAPLE system, we have shown that the two sets of constraints are equivalent. They can be included in the scheme for finding planes.

A second method (also presented by [3]) uses only one plane Π. Let M be a space point which do not lie on Π, m and m' its two images by the two cameras. Let M_1, the point of Π which has the same image by the first camera as M. Using the homography h_Π, we obtain the image of M_1 by the second camera, which is $h_\Pi(m)$. The line $\langle m', h_\Pi(m) \rangle$ is an epipolar line. Two epipolar lines are sufficient to compute the epipole, thus the minimum number of correspondences that are needed in this case is six: four coplanar points, and two points that do not lie in the plane, which is consistent with the mininal number of points needed for the computation of the fundamental matrix using two planes when two of the points belong to both of the planes. Note that if the additional piece of information, coplanarity, was not available, six correspondences would be compatible with an infinity of F-matrices.

Experimental results We first show some experiments conducted with the single plane method. We have first done statistical simulations, using 100 trials. We generated each time a different configuration of 3D points, of which 30 points were taken in a plane, and 20 points off it. The displacements between the camers where varied as well. From there, noisy correspondences were obtained as previously. In each trial, we first computed the homography from images of the planar points, assuming that the segmentation is known. Then we compute an epipolar line for each of the off-plane points. The epipole is obtained as the point e which minimizes the sum of the distances to the epipolar lines:

$$\sum_i d^2(\mathbf{e}, l_i)$$

The non-linear minimization is started from a closed-form solution obtained with two off-planes points which were picked up randomly.

The results were extremely sensitive to noise, as shown by table 4.4, which shows the mean of relative errors on the coordinates of the epipoles. This basic algorithm is very sensitive to noise because it is difficult to localize the intersection of noisy lines, when they all go through

	0.4 pixel	0.8 pixel	1.2 pixel	1.6 pixel
error	0.4864	0.6101	0.6846	0.7287

Table 3: Statistical results of the single plane method.

a fixed part of the retinal plane, which is the image. If the error near the epipole is comparable to the dimension of the image, then almost any point can be obtained. We can see on the top of figure 17 that the intersecting point (430,368) is computed with a big uncertainty, which is to be compared to the noiseless case where (left bottom) the epipole is found at (640,213). Though, we can see that the noisy epipolar lines obtained with the homography are very coherent *in the image*, as they lie very close to the corresponding points. The method might benefit from a more refined technique to find the intersection, such as used recently by [36] in a rather similar approach.

We now turn to a comparison of the multiplane method and the general method. We have used a statistical method, as in 4.3. 50 to 100 points where placed randomly in a controlled number of planes. For each configuration, we have run the general algorithm as well as the multi-plane algorithm. The results are in table 4.4. We can conclude that:

- as soon as the points are in more than three planes, the general method will give better results.
- if the number of planes increases, the results of the general method will improve significantly, whereas the performance of the multiplane method will decrease.
- the multiplane method can be interesting with two planes, and only with two planes.

		number of planes						
image noise	method	2	3	4	5	6	7	8
0.2 pixel	**PLANES**	0.1961	0.2419	0.3056	0.3145	0.2613	0.2996	0.3146
	GENERAL	0.2508	0.1713	0.1158	0.0943	0.0697	0.0702	0.0683
1 pixel	**PLANES**	0.4199	0.4462	0.5955	0.5331	0.5668	0.5806	0.5893
	GENERAL	0.4853	0.4017	0.3390	0.2926	0.3004	0.2749	0.2651

Table 4: Results of the general method, vs results of the multiplane method.

We can explain these results using a simplified error analysis. In the multiplane method, the fundamental matrix is obtained from equations (22). These equations have a form which is similar to equations (2). In order to determine $\mathbf{F}$ from equations (22) with the same precision than from equations (2), the coefficients of (22), which are the coefficients of homography matrices have to be determined with the same precision than the coefficients of (2), which are very simple functions of the pixel coordinates of the points. Let us suppose that the noise has standard deviation $\sigma = 1$ pixel. The coefficients of (2) can be written $C_1 = ab$ or $C_2 = a$, a and b being the affine coordinates of the points. With classical hypothesis, we come to covariances $\sigma_{C_1} = \sigma\sqrt{a^2 + b^2}$, $\sigma_{C_2} = \sigma$. When using 512×512 images, the mean relative error for C_1 (respectively C_2) is thus $\frac{\sqrt{2}}{255} = 5.5\%$ (respectively $\frac{1}{255} = 3.9\%$). According to figure 13, we can estimate an homography matrix only with an average relative error of 13%. In the experiment, the total number of points is fixed, thus when we increase the number of planes, we decrease the number of points per plane, which decreases the precision on the computation of the homography. We obtain more equations, but they are more noisy, and globally the result is worse.

We end by giving a real example of pairs of images where the multi-plane method gives better results than the general method: the points used are taken on the biplane calibration

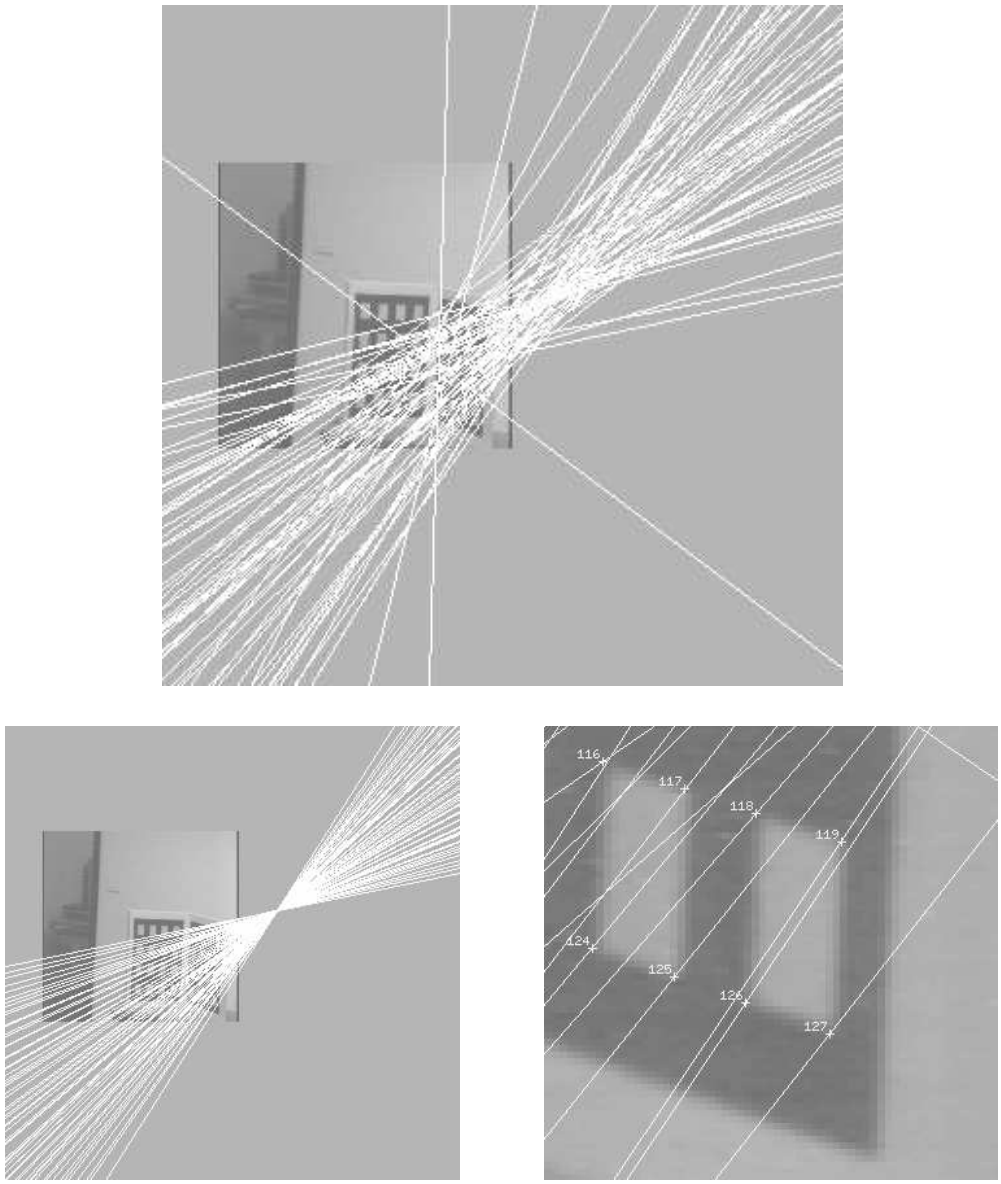

Figure 17: Set of epipolar lines obtained with an homography.

1 2 3 4

Figure 18: The four grid images.

grid. Using a standard calibration procedure, we can obtain reference values for the epipoles. As there is a great number of points in each plane (64) the homography can be determined with a good accuracy. But as all the points lie in a small number of planes (2), and as they are cluttered in a small part of space, the general method will not give localize the epipoles consistently. The table 4.4 shows results obtained with the three displacements between the four images shown on figure 18. We can see a drastic improvement during the first displacement whereas the results remain very similar for the two other displacements.

	$1 \rightarrow 2$			
EXACT	780.08	-427.04	1823.82	-1255.57
GENERAL	-31.68	709.76	174.26	543.61
PLANES	781.70	-455.56	1870.00	-1362.63
	$2 \rightarrow 3$			
EXACT	1634.30	-1910.35	3859.44	-3830.30
GENERAL	744.03	-573.65	1112.87	-950.32
PLANES	943.96	-727.65	1386.38	-1028.40
	$3 \rightarrow 4$			
EXACT	517.49	203.88	349.39	208.09
GENERAL	523.04	205.67	353.92	209.81
PLANES	538.12	206.70	366.34	210.94

Table 5: Results obtained on the grid images.

4.5 Summary

It is known, in the framework of motion analysis that when the points lie in a plane, the general methods will fail, and that some specific methods can be used [74] [38] [17]. We have shown that the situation is similar for the computation of the Fundamental matrix. We have established a very simple and important relation between the homography matrices obtained from the observation of planar surfaces and the Fundamental matrix. Using simulations and real images to validate our analysis, we have shown that the general methods to compute the fundamental matrix are unstable when the points lie near planes. We have then proposed new algorithms to exploit this situation, and have compared their performance to the performance of the general algorithm to determine their domain of applicability.

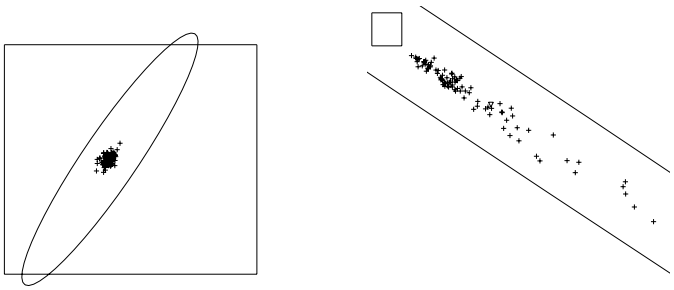

Figure 19: Uncertainty ellipses and noisy epipoles, left:first motion, right:second motion.

5 A stability analysis

5.1 A probabilistic characterization

A classic characterization of uncertainty is to use covariance matrices. If the measurements are modeled by the random vector $\mathbf{x}$, of $\mathbb{R}^{\mathbf{P}}$ of mean $\mathbf{x}_0$ and of covariance $\mathbf{\Lambda_X} = E((\mathbf{x}-\mathbf{x}_0)^T(\mathbf{x}-\mathbf{x}_0))$, then the vector $\mathbf{y} = f(\mathbf{x})$ is a random vector whose first and second order moments can be expressed very simply, up to a first order approximation, as functions of the first and second order moments of $\mathbf{x}$. In effect, the mean is $f(\mathbf{x}_0)$ and the covariance matrix:

$$\mathbf{\Lambda_Y} = \mathbf{J}_f(\mathbf{x}_0)\mathbf{\Lambda_X}\mathbf{J}_f(\mathbf{x}_0)^T \tag{24}$$

Where $\mathbf{J}_f(\mathbf{x}_0)$ is the Jacobian matrix of f, at the point $\mathbf{x}_0$. In our case, the function f associates to the coordinates of the point correspondences the entries of the fundamental matrices eventually found. In the case of a linear criterion, already studied in [79] and [58] (for the computationally identical case of the essential matrix computed from the eight point algorithm), we have an explicit formula for the function f. A different approach is needed to cope with the case of a nonlinear criterion, since we do not have an explicit expression for f. We only know that f minimizes a known criterion, and this can be dealt with using a method based on the implicit functions theorem, presented in [14], and used for instance in [77]. Two examples, one with epipoles near the image center, the other with epipoles far away, are given in figure 19, where we have superimposed the uncertainty ellipses corresponding to a 90% probability, computed from the exact point coordinates, and the image frames. The prediction is accurate, for the order of magnitude, as well as for the major direction of uncertainty. This example illustrates the huge difference of stability which might result from different camera relative positions, and hence the need for a stability analysis. The epipole is nothing else than the projection of the optical center of the other camera. Areas which are projected down onto the image plane in regions far from the image center are more distorted, as seen clearly in this example. This explains why configurations with epipoles which lie far from the image center are less stable, a fact confirmed statistically in section 5.3.

A statistical test has then been performed using 200 configurations of points obtained by variation of cameras and 3D points. The correlation diagram between actual standard deviations (computed over 20 trials for each configuration) and predicted covariances (both from the exact point correspondences: light dots, and from the noisy point correspondences: dark dots), presented figure 20 shows that the correlation between the prediction and the actual covariances is quite high, even in case of prediction from the noisy data.

5.2 Ambiguity and the critical surface

Critical surfaces were known from the photogrammeters of the beginning of the century, who called them "*gefährliche Flächen*". They were then rediscovered and studied theoretically by

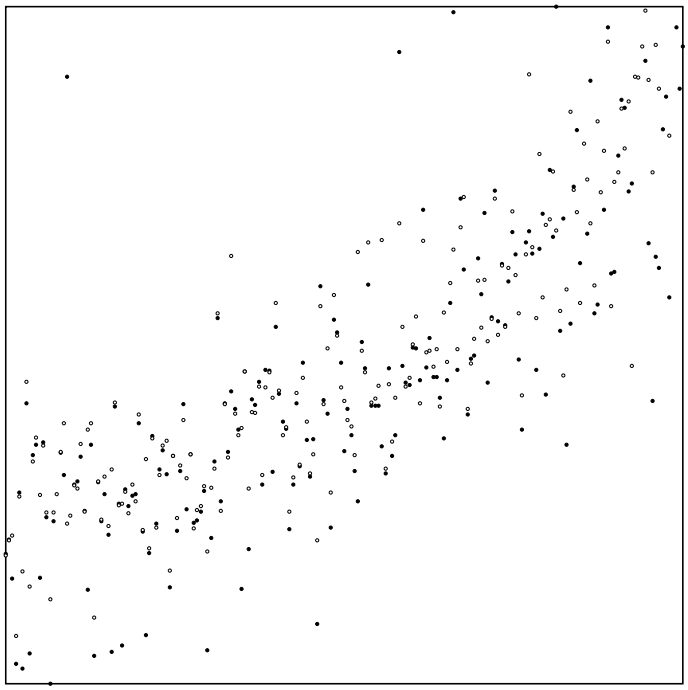

Figure 20: Correlation between computed covariances and predicted covariances.

computer vision scientists in the case of reconstruction from optical flow [48] and point correspondences [41, 50, 28]. We are going to point out some *practical* consequences of the existence of such surfaces. Our approach is to provide algorithms which start from the data which is available to us in uncalibrated images, that is a set of point correspondences between two images. These algorithms provide us a practical means to quantify the proximity of the 3D points which have given rise to point correspondences, to such a critical surface, much the same way than the computation of an homography in Section 4 allowed us to assess the proximity of the 3D points to a plane.

The critical surface and quadratic transforms If all the observed points are in some special configuration, then the problem to obtain fundamental matrices from point correspondences may not have a unique solution, even with an arbitrarily large number of such correspondences. This happens when the measured points lie on some special surfaces called *critical surfaces* and yields several fundamental matrices compatible with the basic constraint : $\mathbf{m}'^T\mathbf{F}\mathbf{m} = 0$. Each of these fundamental matrices gives rise to a displacement which produces identical pairs of views, called *ambiguous*. More precisely, it is not possible to distinguish between the image of the set of 3D points $\mathcal{Q}_1$ observed during displacement $\mathbf{R}_1$, $\mathbf{t}_1$, and the image of a set of points set of 3D points $\mathcal{Q}_2$ observed during displacement $\mathbf{R}_2$, $\mathbf{t}_2$, as illustrated in figure 21. It has been shown [48] that the critical surfaces $\mathcal{Q}_1$ and $\mathcal{Q}_2$ are space quadrics containing the optical centers and the baseline of equations:

$$(\mathbf{R}_1\mathbf{M} + \mathbf{t}_1)^T\mathbf{E}_2\mathbf{M} = 0 \tag{25}$$

$$(\mathbf{R}_2\mathbf{M} + \mathbf{t}_2)^T\mathbf{E}_1\mathbf{M} = 0 \tag{26}$$

It is known that the maximum number of ambiguous fundamental matrices is three [41].

Let us now characterize critical surfaces in terms of image quantities. Given two ambiguous images there exist two fundamental matrices $\mathbf{F}_1$ and $\mathbf{F}_2$ such that for each pair (m, m') of corresponding points,

$$\mathbf{m}'^T\mathbf{F}_1\mathbf{m} = 0 \quad \text{and} \quad \mathbf{m}'^T\mathbf{F}_2\mathbf{m} = 0$$

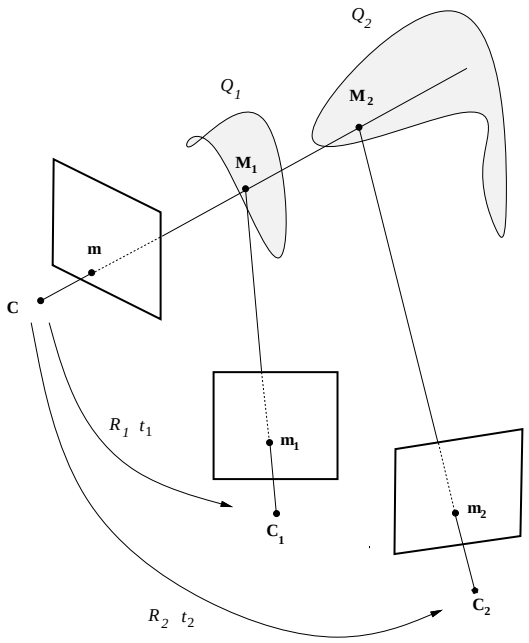

Figure 21: Critical surfaces.

we can conclude from these two equations that:

$$\mathbf{m}' = \mathbf{F}_1\mathbf{m} \times \mathbf{F}_2\mathbf{m} \tag{27}$$

This equation defines in general a *quadratic transformation* between the coordinates of the points in the two images. This is a generalization of the homography which we encountered and studied in the case of planes. Some details about quadratic transforms and their parameterizations can be found in Appendix A. The quadratic transformation allows us to check if image points are close to the projection of a critical surface, much the same way as the homography allowed us to check if they were close to the projection of a plane. The epipoles of the three different fundamental matrices which are solutions to the problem, in an ambiguous situation, are the fundamental points of the quadratic transformation.

Computing the quadratic transformation A first approach is to estimate the 14 parameters of the most general quadratic transformation Φ. The method is also similar to the estimation of the fundamental matrix: it consists in a combination of the linear solution and non-linear minimization with an appropriate parameterization and symmetric Euclidean distance. The non-linear criterion[2] that we minimize is:

$$\min_{\{g_{jx}, g_{jy}, g'_{jx}, g'_{jy}\}\ j=1,2,3, \rho'_1/\rho'_3, \rho'_2/\rho'_3} \sum_i \{d^2(\mathbf{m}'_i, \Phi(\mathbf{m}_i)) + d^2(\mathbf{m}_i, \Phi^{-1}(\mathbf{m}'_i))\} \tag{28}$$

where g_{jx}, g_{jy} (resp g'_{jx}, g'_{jy}) are the coordinates of the direct (resp inverse) fundamental points, which are also the epipoles of the ambiguous fundamental matrices, and the ρ'_i are the scale factors defined in Appendix A.

[2]The fundamental points are supposed to be at finite distance, and we take for example $\mathbf{g}_4 = \mathbf{g}'_4 = \mathbf{i}_4 = (1,1,1)^T$

Let us now describe the linear method used to initialize the search. Writing that Φ is a polynomial homogeneous transformation of the projective plane of degree 2, gives:

$$\left\{\begin{array}{l}\Phi(x_1,x_2,x_3)_1 = a^1_{11}x_1^2 + a^1_{22}x_2^2 + a^1_{33}x_3^2 + a^1_{12}x_1x_2 + a^1_{23}x_2x_3 + a^1_{31}x_3x_1 \\ \Phi(x_1,x_2,x_3)_2 = a^2_{11}x_1^2 + a^2_{22}x_2^2 + a^2_{33}x_3^2 + a^2_{12}x_1x_2 + a^2_{23}x_2x_3 + a^2_{31}x_3x_1 \\ \Phi(x_1,x_2,x_3)_3 = a^3_{11}x_1^2 + a^3_{22}x_2^2 + a^3_{33}x_3^2 + a^3_{12}x_1x_2 + a^3_{23}x_2x_3 + a^3_{31}x_3x_1\end{array}\right. \tag{29}$$

The equations (29) are linear in the 18 entries a^k_{ij}, thus if we have more than 9 point matches, these entries can be obtained by a linear least-squares procedure. Once they are computed, the direct fundamental points are obtained as an approximate solution to the system of equations:

$$\left\{\begin{array}{l}\Phi(x_1,x_2,x_3)_1 = 0 \\ \Phi(x_1,x_2,x_3)_2 = 0 \\ \Phi(x_1,x_2,x_3)_3 = 0\end{array}\right. \tag{30}$$

By exchanging the two images, the inverse fundamental points are computed. We have then to match direct and inverse fundamental points, according to (35), and to compute the scale factors ρ'_1, ρ'_2, ρ'_3. For each of the six distinct configurations (permutations are not to be considered), we solve the least squares problem associated to the 18 linear homogeneous equations with unknowns ρ'_1, ρ'_2, ρ'_3, obtained by identifying (36) with (29). We then select the configuration which gives the smallest value for the criterion (28). We have found that the computation of the 14 parameters of the quadratic transformation is rather unstable. A way to reduce the dimensionality of the problem is to use a fundamental matrix which has already been found. The idea is to start from a fundamental matrix $\mathbf{F}_1$ and to compute a second fundamental matrix $\mathbf{F}_2$ such that $\mathbf{F}_1$ and $\mathbf{F}_2$ define a quadratic transformation (27) minimizing the criterion:

$$\min_{\mathbf{F}_2} \sum_i \{d^2(\mathbf{m}'_i, \mathbf{F}_1\mathbf{m}_i \times \mathbf{F}_2\mathbf{m}_i) + d^2(\mathbf{m}_i, \mathbf{F}_1^T\mathbf{m}'_i \times \mathbf{F}_2^T\mathbf{m}'_i)\} \tag{31}$$

We notice that the expression $\mathbf{F}_1\mathbf{m} \times \mathbf{F}_2\mathbf{m}$ is linear in $\mathbf{F}_2$ which provides a way to initialize the non-linear minimization. This procedure is very comparable to the one used to compute the fundamental matrix studied in Section 3, and the same parameterizations are used for $\mathbf{F}_2$. This method has turned out to be much more efficient than the one consisting in estimating the 14 parameters. Furthermore, the results can be used as a starting point for the more complicated method. We will use the image distances appearing in the criterion (31) as measures of the distance of the three-dimensional points to the corresponding critical surface, and call it *reprojected distance*. Although, given a *fixed* fundamental matrix $\mathbf{F}_1$, the minimization of (31) gives us only an *upper bound* of the reprojected distances of the points (m, m') to the critical surface, this is sufficient for our purposes, as will be seen later. Some experimental comparison of the different computation methods can be found in [42].

Theoretical link between ambiguity and unstability Critical surfaces have been presented in (5.2) as sets of points yielding ambiguous interpretations of motion. Maybank [50] has shown that a configuration whose 3D reconstruction is unstable is close to a critical surface. We are going to provide evidence for the reciprocal property.

The unstability is very clear in the formulation of Horn [28] which defines critical surfaces as sets of points M for which the variation of $\mathbf{m}'^T\mathbf{E}\mathbf{m}$ is a second-order (quadratic) function of the parameters $\mathbf{r}$, $\mathbf{t}$. While the equation he obtains is quite different from (26), he finds properties similar to the one which are described by Maybank [49]. We are going to see that the two forms are indeed equivalent, which will prove that an ambiguous situation is also unstable.

Normalized coordinates are used, the optical center C being mapped onto the optical center C' by the displacement $\mathbf{R}$, $\mathbf{t}$, perturbed by the infinitesimal vectors $\delta\mathbf{r}$, $\delta\mathbf{t}$. The difference of residual values of the Longuet-Higgins equation for unperturbed and perturbed displacement can be expressed in the final coordinate system, using triple products, as:

$$\Delta = [(\mathbf{t} + \delta\mathbf{t}), \mathbf{C}'\mathbf{M}, \mathbf{C}\mathbf{M} + \delta\mathbf{r} \times \mathbf{C}\mathbf{M}] - [\mathbf{t}, \mathbf{C}'\mathbf{M}, \mathbf{C}\mathbf{M}] \tag{32}$$

We have used the fact that an infinitesimal rotation $\delta\mathbf{R}$ can be expressed from $\delta\mathbf{r}$ using the Rodrigues formula, with an infinitesimal $\theta = \|\delta\mathbf{r}\|$:

$$\delta\mathbf{R} = e^{[\delta\mathbf{r}]_\times} = \mathbf{I} + \frac{\sin\theta}{\theta}[\delta\mathbf{r}]_\times + \frac{1-\cos\theta}{\theta}[\delta\mathbf{r}]_\times^2 \sim \mathbf{I} + [\delta\mathbf{r}]_\times$$

The difference Δ in (32) is normally a first order quantity, and the unstable situations are those for which it is a higher order quantity. If we drop the second order term $[\delta\mathbf{t}, \mathbf{C}'\mathbf{M}, \delta\mathbf{r} \times \mathbf{C}\mathbf{M}]$ we obtain by expanding the products:

$$\Delta = [\mathbf{t}, \mathbf{C}'\mathbf{M}, \delta\mathbf{r} \times \mathbf{C}\mathbf{M}] + [\delta\mathbf{t}, \mathbf{C}'\mathbf{M}, \mathbf{C}\mathbf{M}]$$

Using $\mathbf{t} = \mathbf{C}'\mathbf{C}$ and some standard properties of the triple product yields (as also found by [7]):

$$\Delta = [(\mathbf{I} + [\delta\mathbf{r}]_\times)\mathbf{C}'\mathbf{M} - \delta\mathbf{r} \times \mathbf{t} + \delta\mathbf{t}, \mathbf{t}, \mathbf{C}'\mathbf{M}]$$

It is easy to see that this is equivalent to Horn's expression. Now using $\mathbf{M}$ in the initial coordinate system, we obtain, by writing that the triple product is zero:

$$((\mathbf{I} + [\delta\mathbf{r}]_\times)\mathbf{R}\mathbf{M} - \delta\mathbf{r} \times \mathbf{t} + \delta\mathbf{t})^T(\mathbf{t} \times \mathbf{R}\mathbf{M}) = 0 \tag{33}$$

A critical surface given by (26), can be written in the initial coordinate system:

$$(\mathbf{R}_2\mathbf{M} + \mathbf{t}_2)^T(\mathbf{t} \times \mathbf{R}\mathbf{M}) = 0$$

which has the form (33), with:

$$\begin{aligned} \mathbf{I} + [\delta\mathbf{r}]_\times &= \mathbf{R}_2\mathbf{R}^{-1} \\ \delta\mathbf{t} &= \mathbf{t}_2 + \delta\mathbf{r} \times \mathbf{t} = \mathbf{t}_2 - \mathbf{t} + \mathbf{R}_2\mathbf{R}^{-1}\mathbf{t} \end{aligned}$$

5.3 Experimental results

The nature of the motion Since the epipoles are a simple function of the camera displacement, we can expect that the stability of the fundamental matrix computation can be related to the stability of motion estimation [79]. We have studied three cases where the results are unstable:

- small translational component,
- translational component parallel to the image plane (the epipoles are far from the image center),
- pure translation (the fundamental matrix is antisymmetric).

In the two last cases, the nature of the (Euclidean) motion can be characterized from fundamental matrices. Note that whereas the two first cases are known in the context of motion analysis ([12, 79, 31, 71, 80], [12, 53, 1, 79, 29, 8]), the last one is specific to our problem, and comes from the fact that the fundamental matrix in the case of a pure translation depends only on two independent parameters, instead of seven.

We present briefly some experimental results in figure 22 and refer the interested reader to [46] for more simulations and qualitative explanations. In the simulations 100 series of entirely different displacements have been used. Within each series, we have successively varied only:

- the norm of the translation (from 1 to 100),
- the angle between the image plane and the direction of translation (from 5^o to 85^o),
- the angle of the rotation (values: 0.001, 0.002, 0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.5)

For each resulting displacement, we have computed the 2×2 covariance matrix on the coordinates of the epipoles, from which we have obtained an estimate of the major axis of the uncertainty ellipsis for a fixed level of probability. A measure of the relative stability is obtained by taking a ratio[3] of these estimates. For each value of the parameter which is being studied, we then compute the mean value and standard deviation of the ratios over the 100 different series of displacement.

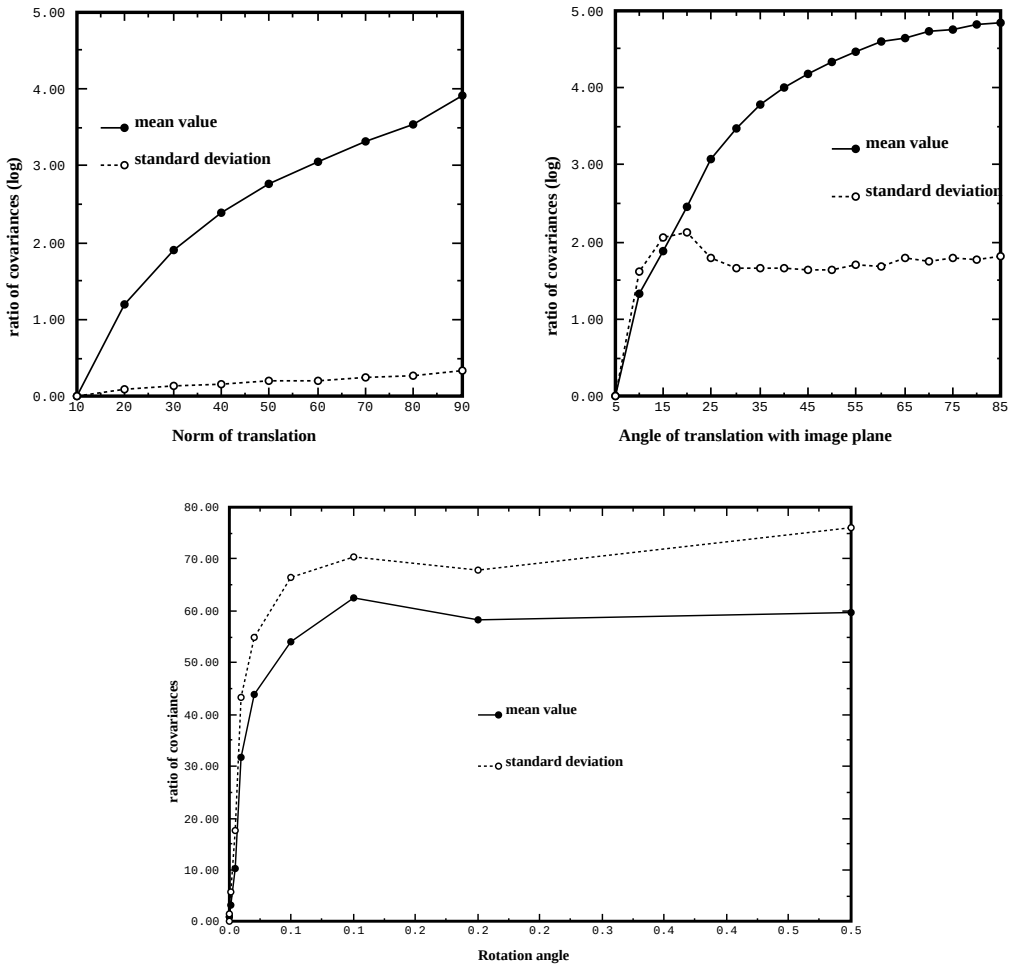

Figure 22: Statistics: the epipole stability as a function of the motion. The higher a point is on the Y-axis, the more stable are the epipoles.

An experiment starting from a critical surface In order to show that critical surfaces are a cause of unstability, we first start from 3D points that are generated on such a surface $\mathcal{Q}$, shown in figure 23. We then construct different sets $\mathcal{S}_d$ of 3D points which lie close to the critical surface. Each point $M_i(d)$ is obtained from the point M_i of the surface $\mathcal{Q}$ from $\mathbf{M}_i \pm d\mathbf{n}_i$, where $\mathbf{n}_i$ is the unit normal to the surface $\mathcal{Q}$ at M_i, and d is a fixed scalar which represents the 3D distance of $\mathcal{S}_d$ to $\mathcal{Q}$. Taking the $\mathcal{S}_d$ instead of the $\mathcal{Q}$ amounts to "add noise to the critical surface", in order to assess the "robustness of unstability", or to evaluate the "critical volume". To assess the stability of fundamental matrix computation, we have then estimated the variance of the coordinates of the epipoles from 50 tries, for different values of the distance to the critical

[3]Since the order of magnitude of these values has quite a large variation in the two first cases, in order to facilitate the visualization of these results on a graph, we have taken a logarithm.

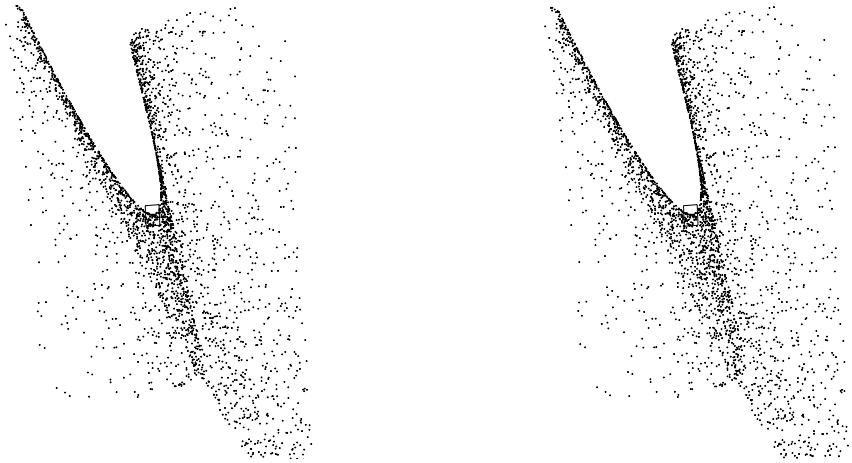

Figure 23: The critical surface used in the experiments (stereogram for cross-fusion).

			$b=0$		$b=0.5$		$b=1$	
d	d_x	d_y	σ_{e_x}	σ_{e_y}	σ_{e_x}	σ_{e_y}	σ_{e_x}	σ_{e_y}
0	0	0	6140	3639	1466	872	1261	788
5	3.89	7.74	10^{-7}	10^{-7}	2935	1765	3749	2305
10	7.60	14.51	10^{-7}	10^{-7}	726.	459	822	492
20	15.19	29.12	10^{-7}	10^{-7}	153	106	280	199
50	89.34	148.53	10^{-7}	10^{-7}	39	40	65	68

Table 6: Influence of the distance to the critical surface and of image noise on the stability.

surface and the image noise. The results appear in table 6, where we also show the mean values d_x and d_y of the retinal disparities between the projections of points of $\mathcal{Q}$ and the projections of the corresponding points of $\mathcal{S}_d$.

Let us comment the results. First, it is clear that the farther the points are from the critical surface, the more stable are the results. When the points are far away from the critical surface, an increase of the image noise increases the covariance of the epipoles, which is to be expected, but when they are very close to the critical surface, the noise induces a reconstruction error which drives the points away from the critical surface, which explains why the variances decrease a little. If there is no image noise, then 3D points are reconstructed exactly. In this case, their 3D distance to the critical surface, even if it is very small, is significant, and unstability does not occur. In the case where there is some image noise, the 3D points are reconstructed with an uncertainty. Now if the original 3D points were close to the critical surface, and if this distance is smaller that the reconstruction uncertainty, then they cannot be distinguished from points lying on the critical surface, and thus unstability will occur. Thus, the volume for which unstability occurs depends on the 2D noise and we call it the *critical volume*.

A global experiment So far, we have always started from synthetic data which was created to illustrate some facts. Now we start from the image data, such that it would be available to an algorithm, and we try to explain the sources of uncertainty. This experiment was carried on using synthetic data because at that time we did not have a reliable system to obtain automatically point matches, but the principle would be exactly the same with real data. In this experiment, we try to account simultaneously for two sources of unstability, the proximity to a critical

displacements (increasing unstability)	critical surface at less than 10 pixels	average distance of epipoles to image center
1-100	9%	754.6 pixels
101-200	13%	1164 pixels
201-300	31%	1783 pixels
301-400	40%	2624 pixels
401-500	49%	25280 pixels

Table 7: Sources of unstability in a statistical experiment.

surface, and the distance of the epipole to the image center. Note that we have eliminated data with small retinal disparity in order to ignore the unstability due to small and pure translations. The image noise is 2 pixels. For each of the 500 displacements, we have computed the epipoles and their covariance matrices, and ordered the trials by increasing unstability. The horizontal axis in figure 24 represents unstability increasing from left to right.

We have first considered the distance of the epipole to the image center, represented (using a non-linear scale) on the Y-axis. There is a correlation between this distance, and unstability, quantified by the leftmost and rightmost columns of table 7.

The next idea is to try to fit a critical surface, by computing the reprojected distance to a critical surface using the method described in Section 5.2. Since the 3D points are chosen randomly, their probability to lie on a critical surface is almost zero. However, and this is one of our findings, they may lie *near* a critical surface, which means that they are in a critical volume. The idea is, after estimating the fundamental matrix $\mathbf{F}_1$ from the point correspondences, to find the fundamental matrix $\mathbf{F}_2$ which minimizes (27). This is like trying to fit a critical surface to the 3D points which have given rise to the point correspondences. If the residual distance, which is the value of the criterion (27) at the minimum, is high, it means that no fit can be found, and thus the critical surface does not exist. But if the residual is low, it means that the 2D points lie near the projection of a critical surface, the distance of the points to the projection of the fitting critical surface being given by the residual. Of course, there is a continuum of possibilities, and we have chosen the threshold of 10 pixels, for which we know that unstability is still significant, as shown by the example presented in table 6.

The black dots in figure 24 are those for which the distance is under the threshold. Let us consider two points N_1 and N_2 in figure 24, with approximately the same horizontal coordinate, but for which the vertical coordinates are different, say $y_1 > y_2$. The points have the same stability, but N_2 correspond to a motion yielding an epipole which is closer to the image center than N_1. The reason may be that N_2 represents a configuration which is close to a critical surface. Now we can notice that these points (the dark dots) are statistically below the light dots (corresponding to distances to the critical surface which are more than 10 pixels), which validate this hypothesis. Another thing which may be observed is that there are more black dots in the area of high unstability (right), as shown in the middle column of table 7 as well as in figure 24. Thus, the combination of the proximity to a critical surface and the direction of translation provides a better explanation for unstability than any of these two causes in isolation.

Another important observation is the omnipresence of the critical surface, which is at less than 10 pixels in 28% of the displacements. Although the critical surfaces do not exist *exactly* in normal scenes in the sense that real objects rarely are critical surface, they have a large practical importance since our experiments show that the *critical volume* where the points have to lie in order to yield some unstability is rather large.

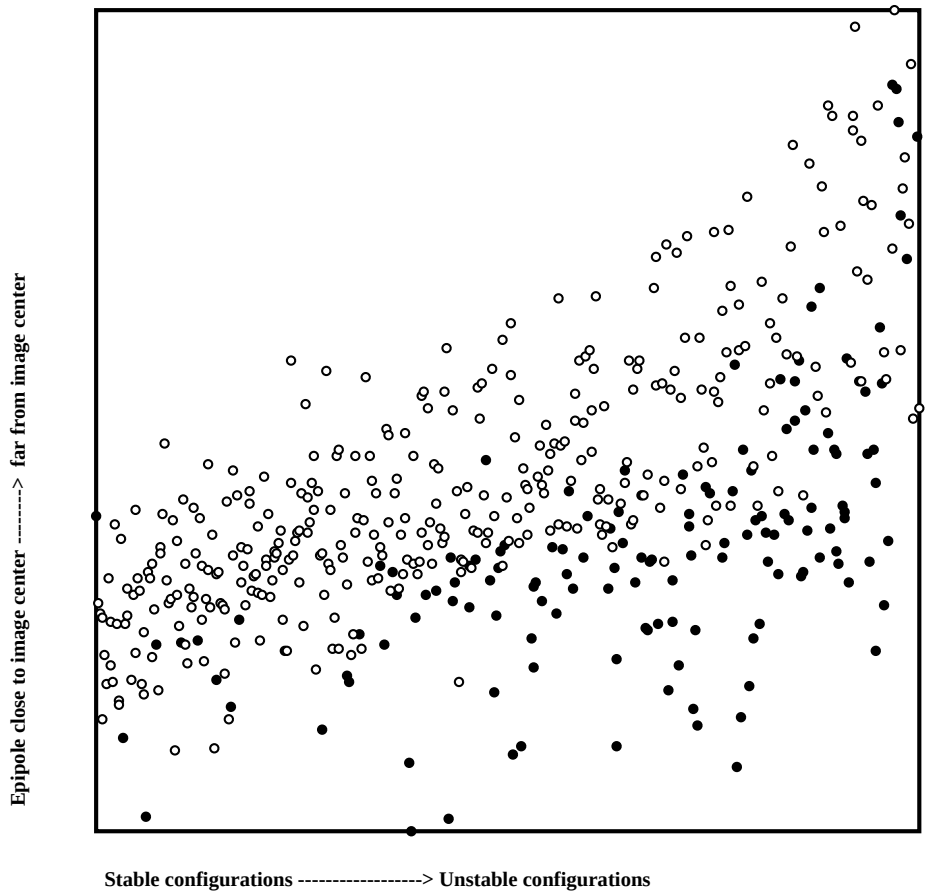

Figure 24: A global experiment to characterize the causes of unstability (see text).

5.4 Summary

In this section, we have studied the influence of the camera motion on the stability of the estimation of the fundamental matrix. Two tools have been introduced, a probabilistic characterization of the stability through the computation of the covariance matrix of the estimate, and a method to compute the reprojected distance to the closest critical surface from image data through the estimation of a quadratic transformation. Using these tools we have been able to characterize the unstable situations. They can arise from the nature of the motion (small translational component, translational component parallel to the image plane, and pure translation), but also in a more subtile way, from the interaction of motion and 3D structure of the scene, which can be described by a critical surface. These characterizations have been validated experimentally through statistical simulations.

6 Conclusion

Three types of transformations attached to the projective structure of a system of two cameras have been considered: the correlation between points and epipolar lines, the homography between images of a plane, and the quadratic transformation between images of quadrics. We have made several links between them, and shown that the two latter ones can be used to characterize the stability of the *fundamental matrix*, which provides an algebraic formulation for the first one. For each of these transformations, we have also provided a computationnal algorithm consisting of both a linear solution and a non-linear solution. The latter one allows to express the constraints of a problem by an appropriate parameterization, and a criterion which is a distance in measurement spaces. In particular, in the case of the fundamental matrix, we have done an analysis and provided several experimental evidence to show that our approach provides much more robust results. Although the basic algorithm for their computation through the fundamental matrix is simple, the epipoles turn out to be rather unstable quantities. One of the contributions of this work is not only to provide computationnal solutions, but also to characterize the unstable situations, which we have found to be numerous. It can be noted that we can have very coherent epipolar lines in the image, whereas the location of the epipole is very inacurrate. Such situations are not a problem for stereo matching, the task for which epipolar geometry has been traditionnaly used, but will prevent stable recovery of projective structures or invariants. This is probably the price to pay for the large generality of the formalism.

The fundamental matrix is indeed central to all problems involving uncalibrated cameras, and two viewpoints. It captures in a very compact manner the epipolar geometry of a stereo rig. This geometry is purely projective and independent of the intrinsic parameters of the cameras. Even though it is true that, by taking them into account, the fundamental matrix can be shown to be equivalent to the essential matrix introduced by Longuet-Higgins [40], we believe and have shown, as well as others ([13, 23, 16, 68, 42, 47] just to cite the major papers), that abstracting the idea of the fundamental matrix has led to deep theoretical insights about the kind of three-dimensional information that can be recovered from sets of images (projective, affine, Euclidean), and to practical ways of incrementally recovering such information from the environment depending upon the task at hand and the computing power available.

This idea is at the basis of a flurry of very recent and promising developments in three-dimensional vision that will most likely lead to robust and flexible (i.e. at last practical) solutions for a whole gamut of applications involving the use of computer vision in simulation, virtual reality, and robotics applications. It is indeed remarkable that so many applications of the theory were studied by several authors in such a short amount of time. To emphasize this point, and answer the question "what is the Fundamental matrix good for ?" we now list a number of recent papers which have studied tasks to be performed when the only information relating the cameras are the Fundamental matrix (or matrices):

- Recovery of the 3D projective structure of a scene from point matches [13, 23, 68, 59], and of the relative affine structure [13, 67, 69],

- Obtention of projective invariants [67, 24, 22],
- Prediction of image features in an image from image features in two other images [2, 54, 9] (positions) [19] (positions,orientations,curvatures),
- Synthesis of an image from several images [35],
- Convex-hull computation and plane positionning [62],
- Segmentation of rigid independent motions [56, 72, 78],
- Stereo analysis: rectification of images [27, 15] and stereo matching with uncalibrated cameras [61] (feature-based) [15] (area-based) [63, 11]. (taking orientation into account)
- Self-calibration of a moving camera [51, 16, 44, 25].

The Fundamental matrix represents indeed the minimal information (two views, no additional hypotheses), in a hierarchy of representations obtained by making further assumptions and adding views [76, 47]. As a consequence, it is a theoretical and practical tool of primary importance. Its knowledge is essential for further use of the hierarchy of representations. For example, if an addition to the Fundamental matrix, a certain 3-D vector representing the plane at infinity is known, then affine structure (considered for instance in [60, 9, 66, 52, 4]) could be recovered. It is to be noted that there are no need for multiple matches across several frames, since the fundamental matrices can be computed independently from pairs of views. A first extension of this work is however to consider multiple views, which leads to complex dependencies between fundamental matrices [47]. A second extension of this work is to consider the case of lines, for which some results have been already obtained [26, 76]. However, correspondences between lines of at least three views are needed. The resulting information which is possible to extract could be obtained more robustly, but at the cost of starting from more data than approaches using only two views and the fundamental matrix, which is the minimal expression of 3D information, and thus the most general.

Acknowledgements

The authors would like to thank R. Deriche and T. Papadopoulo, who contributed to parts of this work, and S. de Ma, S. Maybank, N. Navab, L. Robert, and Z. Zhang with whom they enjoyed enlightening discussions. The detailed comments of an anonymous reviewer were also most useful to help improve details of the presentation.

A Quadratic transformations and their parameterizations

Quadratic transformations are mappings of $\mathcal{P}^2$ into $\mathcal{P}^2$, whose coordinates are homogeneous polynomials of degree 2, which are invertible, and whose inverse are also homogeneous polynomials of degree 2. The most simple example is the reciprocal transformation, defined by:

$$\Phi_0(\mathbf{x}) = (x_2x_3, x_3x_1, x_1x_2)^T$$

From this definition, we can see that Φ_0 is defined in each point of $\mathcal{P}^2$, except for the points $\mathbf{i}_1 = (1,0,0)^T$, $\mathbf{i}_2 = (0,1,0)^T$ and $\mathbf{i}_3 = (0,0,1)^T$, which are called fundamental points of Φ_0. We also notice that Φ_0 is invertible, since it is it own inverse.

In the general case, a quadratic transformation Φ has also three fundamental points $\mathbf{g}_1$, $\mathbf{g}_2$, $\mathbf{g}_3$ which are distinct of those of Φ^{-1}, $\mathbf{g}'_1$, $\mathbf{g}'_2$, $\mathbf{g}'_3$. and we have:

$$\Phi = \mathbf{A}\Phi_0\mathbf{B} \tag{34}$$

where $\mathbf{A}$ and $\mathbf{B}$ are two collineations which can interpreted as changes of retinal coordinates:

$$\begin{array}{cccc} \mathbf{A}\mathbf{i}_1 = \mathbf{g}'_1 & \mathbf{A}\mathbf{i}_2 = \mathbf{g}'_2 & \mathbf{A}\mathbf{i}_3 = \mathbf{g}'_3 & \mathbf{A}\mathbf{i}_4 = \mathbf{g}'_4 \\ \mathbf{B}\mathbf{g}_1 = \mathbf{i}_1 & \mathbf{B}\mathbf{g}_2 = \mathbf{i}_2 & \mathbf{B}\mathbf{g}_3 = \mathbf{i}_3 & \mathbf{B}\mathbf{g}_4 = \mathbf{i}_4 \end{array} \tag{35}$$

where $\mathbf{i}_4 = (1,1,1)^T$. The inverse of Φ is $\Phi^{-1} = \mathbf{B}^{-1}\Phi_0\mathbf{A}^{-1}$. The point $\mathbf{g}_4$ can be chosen arbitrarily, whereas the point $\mathbf{g}'_4$ is determined by Φ [65]. Thus $\mathbf{A}$ depends on 8 parameters (the projective coordinates of the points $\mathbf{g}'_i$, $i = 1,2,3,4$) and $\mathbf{B}$ depends on 6 parameters (the projective coordinates of the points $\mathbf{g}_i$, $i = 1,2,3$). Thus Φ depends on 14 parameters, which is consistent with (27), where Φ is defined by two fundamental matrices, which gives 7+7 parameters. In relation (34), the matrices $\mathbf{A}$ and $\mathbf{B}$ do not play a symmetrical role, but this relation can be rewritten in a a more symmetrical way. Let us apply the formula for changing projective coordinates (see for example [14]):

$$\mathbf{A} = [\mathbf{g}'_1\ \mathbf{g}'_2\ \mathbf{g}'_3] \underbrace{\begin{bmatrix} \rho'_1 & 0 & 0 \\ 0 & \rho'_2 & 0 \\ 0 & 0 & \rho'_3 \end{bmatrix}}_{\mathbf{D}}$$

Since $[\mathbf{g}'_1\ \mathbf{g}'_2\ \mathbf{g}'_3]$ is invertible, there is a bijective relation between $\mathbf{g}'_4$ and $(\rho'_1, \rho'_2, \rho'_3)$, thus we can rewrite (34) as:

$$\Phi = \mathbf{A}_1\Phi_1\mathbf{B}_1 \tag{36}$$

where $\mathbf{A}_1$ and $\mathbf{B}_1$ are obtained from (35) with an arbitrary choice of $\mathbf{g}_4$ *and* $\mathbf{g}'_4$ and:

$$\Phi_1(\mathbf{x}) = \mathbf{D}\Phi_0(\mathbf{x}) = (\rho'_1 x_2 x_3, \rho'_2 x_3 x_1, \rho'_3 x_1 x_2)^T$$

Thus we have as the 14 parameters twelve coordinates for fundamental points (direct and reverse) and two scale ratios.

References

[1] G. Adiv. Inherent ambiguities in recovering 3-D motion and structure from a noisy flow field. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:477–489, 1989.

[2] E.B. Barrett, M.H. Brill, N.N. Haag, and P. M. Payton. Invariant linear methods in photogrammetry and model-matching. In J.L. Mundy and A. ¡Zisserman, editors, *Geometric invariance in computer vision*, chapter 14, pages 277–292. MIT Press, 1992.

[3] P.A. Beardsley. *Applications of projective geometry to computer vision.* PhD thesis, University of Oxford, 1992.

[4] P.A. Beardsley, A. Zisserman, and D.W. Murray. Navigation using affine structure from motion. In *Proc. European Conference on Computer Vision*, pages 85–96, Stockholm, Sweden, 1994.

[5] F. L. Bookstein. Fitting conic sections to scattered data. *Computer Graphics and Image Processing*, 9(1):56–71, Jan 1979.

[6] H.S.M. Coxeter. *Projective Geometry.* Springer Verlag, second edition, 1987.

[7] K. Daniilidis. *Zur Fehlerempfindlichkeit in der Ermittlung von Objektbeschreibungen und relativen Bewegugen aus monokularen Bildfolgen.* PhD thesis, University of Karsruhe, 1992.

[8] K. Daniilidis and H.-H. Nagel. Analytical results on error sensitivity of motion estimation from two views. *Image and Vision Computing*, 8:297–303, 1990.

[9] S. Demey, A. Zisserman, and P.A. Beardsley. Affine and projective structure from motion. In *Proc. British Machine Vision Conference*, pages 49–58, Leeds, UK, 1992.

[10] R. Deriche, Z. Zhang, Q.-T. Luong, and O.D. Faugeras. Robust recovery of the epipolar geometry for an uncalibrated stereo rig. In *Proc. European Conference on Computer Vision*, pages 567–576, Stockholm, Sweden, 1994.

[11] F. Devernay and O.D. Faugeras. Computing differential properties of 3-D shapes from stereoscopic images without 3-D models. In *cvpr*, pages 208–213, Seattle, WA, 1994.

[12] J.Q. Fang and T.S. Huang. Some experiments on estimating the 3D motion parameters of a rigid body from two consecutive image frames. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:545–554, 1984.

[13] O.D. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig. In *Proc. European Conference on Computer Vision*, pages 563–578, 1992.

[14] O.D. Faugeras. *Three-dimensional computer vision: a geometric viewpoint.* MIT Press, 1993.

[15] O.D. Faugeras, B. Hotz, and al. Real time correlation-based stereo: algorithm, implementations and applications. *The International Journal of Computer Vision*, 1994. To appear.

[16] O.D. Faugeras, Q.-T. Luong, and S.J. Maybank. Camera self-calibration: theory and experiments. In *Proc. European Conference on Computer Vision*, pages 321–334, Santa-Margerita, Italy, 1992.

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

[18] O.D. Faugeras, F. Lustman, and G. Toscani. Motion and Structure from point and line matches. In *Proc. International Conference on Computer Vision*, pages 25–34, June 1987.

[19] O.D. Faugeras and L. Robert. What can two images tell us about a third one ? Technical Report RR-2018, INRIA, 1993. To appear in IJCV.

[20] O.D. Faugeras and G. Toscani. The calibration problem for stereo. In *Proceedings of CVPR'86*, pages 15–20, 1986.

[21] L.E. Garner. *An outline of projective geometry.* Elsevier North Holland, 1981.

[22] P. Gros and L. Quan. 3d projective invariants from two images. In *Geometric methods in computer vision II, SPIE optical instrumentation and applied science*, San Diego, 1993.

[23] R.I. Hartley. Estimation of relative camera positions for uncalibrated cameras. In *Proc. European Conference on Computer Vision*, pages 579–587, 1992.

[24] R.I. Hartley. Cheirality invariants. In *Proc. DARPA Image Understanding Workshop*, pages 745–753, U. of Maryland, 1993.

[25] R.I. Hartley. An algorithm for self calibration from several views. In *Proc. Conference on Computer Vision and Pattern Recognition*, pages 908–912, Seattle, WA, 1994.

[26] R.I. Hartley. Projective reconstruction from line correspondences. In *Proc. Conference on Computer Vision and Pattern Recognition*, pages 903–907, Seattle, WA, 1994.

[27] R.I. Hartley and R. Gupta. Computing matched-epipolar projections. In *Proc. Conference on Computer Vision and Pattern Recognition*, pages 549–555, New-York, 1993.

[28] B.K.P. Horn. Relative orientation. *The International Journal of Computer Vision*, 4(1):59–78, Jan. 1990.

[29] B.K.P. Horn and E.J. Weldon. Direct methods for recovering motion. *The International Journal of Computer Vision*, 2(1):51–76, 1988.

[30] T.S. Huang and O.D. Faugeras. Some properties of the E-matrix in two view motion estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:1310–1312, 1989.

[31] C.P. Jerian and R. Jain. Structure from motion. a critical analysis of methods. *IEEE Transactions on Systems, Man and Cybernetics*, 21(3):572–587, 1991.

[32] D.G. Jones and J. Malik. A Computational Framework for Determining Stereo Correspondence from a Set of Linear Spatial Filters. In *Proc. European Conference on Computer Vision*, pages 395–410, 1992.

[33] K. Kanatani. Computational projective geometry. *Computer Vision, Graphics, and Image Processing. Image Understanding*, 54(3), 1991.

[34] K. Kanatani. *Geometric computation for machine vision*. Oxford university press, 1992.

[35] S. Laveau and O.D. Faugeras. 3-D scene representation as a collection of images. In *Proc. International Conference on Pattern Recognition*, Jerusalem, Israel, 1994. To appear.

[36] J.M. Lawn and R. Cipolla. Robust egomotion estimation from affine motion parallax. Technical Report CUED/F-INFENG/TR 160, University of cambridge, Jan 1994. a shorter version appeared at ECCV'94.

[37] C.H. Lee. Time-varying images: the effect of finite resolution on uniqueness. *Computer Vision, Graphics, and Image Processing. Image Understanding*, 54(3):325–332, 1991.

[38] C. Longuet-Higgins. The reconstruction of a scene from two projections: configurations that defeat the 8-point algorithm. In *Proc. 1st Conf. on Artificial intelligence applications*, pages 395–397, Denver, 1984.

[39] H. C. Longuet-Higgins and K. Prazdny. The interpretation of moving retinal images. *Proceedings of the Royal Society of London*, B 208:385–387, 1980.

[40] H.C. Longuet-Higgins. A Computer Algorithm for Reconstructing a Scene from Two Projections. *Nature*, 293:133–135, 1981.

[41] H.C. Longuet-Higgins. Multiple interpretations of a pair of images of a surface. *Proc. of the Royal Society London A*, 418:1–15, 1988.

[42] Q.-T. Luong. *Matrice fondamentale et auto-calibration en vision par ordinateur*. PhD thesis, Universite de Paris-Sud, Orsay, Dec. 1992.

[43] Q.-T. Luong, R. Deriche, O.D. Faugeras, and T. Papadopoulo. On determining the Fundamental matrix: analysis of different methods and experimental results. Technical Report RR-1894, INRIA, 1993. A shorter version appeared in the Israelian Conf. on Artificial Intelligence and Computer Vision.

[44] Q.-T. Luong and O.D. Faugeras. Camera calibration, scene motion and structure recovery from point correspondences and fundamental matrices. Submitted to IJCV, 1993.

[45] Q.-T. Luong and O.D. Faugeras. Determining the Fundamental matrix with planes: unstability and new algorithms. In *Proc. Conference on Computer Vision and Pattern Recognition*, pages 489–494, New-York, 1993.

[46] Q.-T. Luong and O.D. Faugeras. Stratifed projective motion analysis: Fundamental matrix and self-calibration. In preparation, 1994.

[47] Q.-T. Luong and T. Viéville. Canonical representations for the geometries of multiple projective views. *CVGIP: image understanding*, 1994. To appear.

[48] S.J. Maybank. The angular velocity associated with the optical flow field arising from motion through a rigid environment. *Proc. of the Royal Society London A*, 401:317–326, 1985.

[49] S.J. Maybank. The projective geometry of ambiguous surfaces. *Proc. of the Royal Society London A*, 332:1–47, 1990.

[50] S.J. Maybank. Properties of essential matrices. *International journal of imaging systems and technology*, 2:380–384, 1990.

[51] S.J. Maybank and O.D. Faugeras. A Theory of Self-Calibration of a Moving Camera. *The International Journal of Computer Vision*, 8(2):123–151, 1992.

[52] P.F. McLauchlan, I.D. Reid, and D.W. Murray. Recursive affine structure and motion from image sequences. In *Proc. European Conference on Computer Vision*, pages 217–224, Stockholm, Sweden, 1 1994.

[53] A. Mitiche, X. Zhuang, and R. Haralick. Interpretation of optical flow by rotation decoupling. In *Proc. IEEE workshop on computer vision*, pages 195–200, Miami Beach, FL, 1987.

[54] J. Mundy, R.P. Welty, M.H. Brill, P.M. Payton, and Barrett. 3-D model aligment without computing pose. In *Proc. DARPA Image Understanding Workshop*, pages 727–735, San Mateo, CA, 1992.

[55] J. L. Mundy and A. Zisserman, editors. *Geometric invariance in computer vision*. MIT Press, 1992.

[56] E. Nishimura, G. Xu, and S. Tsuji. Motion segmentation and correspondence using epipolar constraint. In *Proc. 1st Asian Conf. Computer Vision*, pages 199–204, Osaka, Japan, 1993.

[57] S.I. Olsen. Epipolar line estimation. In *Proc. European Conference on Computer Vision*, pages 307–311, 1992.

[58] J. Philip. Estimation of three-dimensional motion of rigid objects from noisy observations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(1):61–66, 1991.

[59] J. Ponce, D.H. Marimont, and T.A. Cass. Analytical methods for uncalibrated stereo and motion reconstruction. In *Proc. European Conference on Computer Vision*, pages 463–470, Stockholm, Sweden, 1994.

[60] L. Quan and R. Mohr. Affine shape representation from motion through reference points. *Journal of mathematical imaging and vision*, 1:145–151, 1992.

[61] L. Robert. *Reconstruction de courbes et de surfaces par vision stéréoscopique. Applications a la robotique mobile*. PhD thesis, Ecole Polytechnique, 1993.

[62] L. Robert and O.D. Faugeras. Relative 3D Positioning and 3D Convex Hull Computation from a Weakly Calibrated Stereo Pair. In *Proc. International Conference on Computer Vision*, pages 540–543, Berlin, Germany, May 1993.

[63] L. Robert and M. Hebert. Deriving orientation cues from stereo images. In *Proc. European Conference on Computer Vision*, pages 377–388, Stockholm, Sweden, 1994.

[64] P.D. Sampson. Fitting conic sections to very scattered data. an iterative refinement of the Bookstein algorithm. *Computer Graphics and Image Processing*, 18(1):97–108, Jan. 1982.

[65] J.G. Semple and G.T. Kneebone. *Algebraic Projective Geometry*. Oxford: Clarendon Press, 1952. Reprinted 1979.

[66] L.S. Shapiro, A. Zisserman, and M. Brady. Motion from point matches using affine epipolar geometry. Technical Report OUEL 1994/93, Oxford University., June 1993. A shorter version appeared at ECCV'94.

[67] A. Shashua. On geometric and algebraic aspects of 3D affine and projective structures from perspective 2D views. In *Proceedings of the 2nd European Workshop on Invariants*, Ponta Delagada, Azores, 1993. Also MIT AI Memo No. 1405, July 1993.

[68] A. Shashua. Projective depth: a geometric invariant for 3d reconstruction from two perspective/orthographic views and for visual recognition. In *Proc. International Conference on Computer Vision*, pages 583–590, Berlin, Germany, 1993.

[69] A. Shashua and N. Navab. Relative affine structure: Theory and application to 3D reconstruction from perspective views. In *Proc. Conference on Computer Vision and Pattern Recognition*, pages 483–489, Seattle, WA, 1994.

[70] D. Sinclair, A. Blake, S. Smith, and C. Rothwell. Planar region detection and motion recovery. In *Proc. British Machine Vision Conf.*, pages 59–68, 1992.

[71] M.E. Spetsakis. A linear algorithm for point and line-based structure from motion. *Computer Vision, Graphics, and Image Processing. Image Understanding*, 56(2):230–241, 1992.

[72] P.H.S. Torr and D.W. Murray. Stochastic motion clustering. In *Proc. European Conference on Computer Vision*, pages 328–337, Stocholm, Sweden, 1994.

[73] R.Y. Tsai. An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision. In *Proceedings CVPR '86, Miami Beach, Florida*, pages 364–374. IEEE, June 1986.

[74] R.Y. Tsai and T.S. Huang. Estimating Three-dimensional motion parameters of a rigid planar patch, II: singular value decomposition. *IEEE Transactions on Acoustic, Speech and Signal Processing*, 30, 1982.

[75] R.Y. Tsai and T.S. Huang. Uniqueness and estimation of three-dimensional motion parameters of rigid objects wirth curved surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:13–27, 1984.

[76] T. Viéville, Q.-T. Luong, and O.D. Faugeras. Motion of points and lines in the uncalibrated case. *Intl. Journal of Computer Vision*, 1994. To appear.

[77] T. Viéville and P. Sander. Using pseudo kalman-filters in the presence of constraints. Technical Report RR-1669, INRIA, 1992.

[78] J. Weber and J. Malik. Rigid body segmentation and shape description from dense optical flow under weak perspective. Dep. of EECS, University of California at Berkeley, 1994.

[79] J. Weng, T.S. Huang, and N. Ahuja. Motion and structure from two perspective views: algorithms, error analysis and error estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(5):451–476, 1989.

[80] Z. Zhang and O.D. Faugeras. *3D dynamic scene analysis*. Springer-Verlag, 1992.