International Journal of Computer Vision 22(3), 261–289 (1997)

Self-Calibration of a Moving Camera from Point Correspondences and Fundamental Matrices*

Q.-T. LUONG
SRI International, 333 Ravenswood av, Menlo Park, CA 94025, USA
luong@ai.sri.com

O.D. FAUGERAS
I.N.R.I.A., 2004 route des Lucioles, B.P. 93 06902 Sophia-Antipolis, France
faugeras@sophia.inria.fr

Received (Bolles Office) 1993; (Kanade Office) December 8, 1995; Accepted October 24, 1995

Abstract. We address the problem of estimating three-dimensional motion, and structure from motion with an uncalibrated moving camera. We show that point correspondences between three images, and the fundamental matrices computed from these point correspondences, are sufficient to recover the internal orientation of the camera (its calibration), the motion parameters, and to compute coherent perspective projection matrices which enable us to reconstruct 3-D structure up to a similarity. In contrast with other methods, no calibration object with a known 3-D shape is needed, and no limitations are put upon the unknown motions to be performed or the parameters to be recovered, as long as they define a projective camera.

The theory of the method, which is based on the constraint that the observed points are part of a static scene, thus allowing us to link the intrinsic parameters and the fundamental matrix via the absolute conic, is first detailed. Several algorithms are then presented, and their performances compared by means of extensive simulations and illustrated by several experiments with real images.

Keywords: camera calibration, projective geometry, Euclidean geometry, Kruppa equations

1. Introduction and Motivations

The problem of estimating the three-dimensional motion of a camera from a number of token matches has received a lot of attention in the last fifteen years. Having detected and matched such tokens as points or lines in two or more images, researchers have developed methods for estimating the three-dimensional camera displacement, assuming a moving camera and a static object. This problem is equivalent to the problem of estimating the three-dimensional motion of an object observed by a static camera. The camera is modeled as a pinhole and its internal parameters are supposed to be known (the pinhole model and the internal parameters are defined later). This is the full perspective case. Other researchers have assumed less general image formation models such as the orthographic model, for example Ullman (1979). In this article we will assume the most general case of the full perspective image formation model.

When matching points, two views are sufficient and the computation of the motion is usually based upon the estimation of a matrix called the Essential, or E-matrix after Longuet-Higgins (1981) who first published a linear algorithm (called the eight-point algorithm because it requires eight point correspondences over two frames) for estimating this matrix and recover

*This work was partially supported by the EC under Esprit grant 5390, Real Time Gaze Control.

the camera displacement from it from a number of point matches. The properties of the E-matrix are now well understood after the work of Faugeras and Maybank (1990), Huang and Faugeras (1989), and Maybank (1990). This matrix must satisfy a number of algebraic constraints which are not taken into account by the eight-point algorithm. Taking these constraints into account forces to use nonlinear methods such as the five-point algorithm of Faugeras and Maybank (1990).

The internal parameters of the cameras are traditionally determined by observing a known calibration object (see Tsai (1989) for a review), prior to the execution of the vision task. However, there are several applications for which a calibration object is not available, or its use is too cumbersome. The thrust of this paper is to extend the previous results to the case where the internal parameters of the camera are unknown, still assuming the full perspective model. Our guiding light will be projective geometry which we found to be extremely useful both from the theoretical point of view in that it has allowed us to express the geometry of the problem in a much simpler way and from the practical point of view in that this formal simplicity can be transported to algorithmic simplicity.

We will show that if we take three snapshots of the environment, each time establishing sufficiently many point correspondences between the three pairs of images, we can a) recover the epipolar geometry of each pair of images b) recover the intrinsic parameters of the camera (which we assume not to be changing during the motion) and c) recover the motion of the camera (using already published algorithms). The focus of the paper is on point b), point a) being described elsewhere.

Section 2 will be dedicated to the geometric and algebraic modeling of the problem and to a description of the relations of the present approach to previous ones. In particular, we will tie the intrinsic parameters to the image of the absolute conic, define the fundamental matrix which is analog to the essential matrix in the uncalibrated case, relate it to the intrinsic parameters. We will also define the Kruppa equations from which we will be able to estimate the intrinsic parameters and relate them to the work on the essential matrix. Section 3 will build upon the theoretical results of Section 2 and describe a method for recovering the intrinsic parameters of the camera and therefore its motion, as detailed in Section 4. Examples with real images are given in Section 5. Finally, in Section 6, we conclude and compare our work to that of others.

2. Background and Theory

In this section we lay the ground for the solution of the problem of estimating the motion of a camera with unknown intrinsic parameters. First we consider the case of a single camera and introduce the camera model and the intrinsic parameters. We make heavy use of simple projective geometry. We show that even for a single camera, projective geometry offers a rich description of the geometry of the problem through the introduction of the absolute conic which is fundamental in motion analysis. We then consider the case of two cameras and describe their geometric relations. We show that these relations can be summarized very simply by the epipolar correspondence (geometric viewpoint) or the fundamental matrix (algebraic viewpoint). We then describe the relationship between the fundamental matrix and the intrinsic parameters of the camera through various complementary approaches.

2.1. The Pinhole Model, Intrinsic and Extrinsic Parameters, the Absolute Conic

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 (see Fig. 1). The optical axis is the line going through C and perpendicular to the retinal plane. It pierces that plane at the principal point c. If we consider an orthonormal system of coordinates in the retinal plane, centered at c, say (c, x_c, y_c) 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_C, Y_C, Z_C). In these two systems of coordinates, the relationship between the coordinates of m, image of M is particularly simple

$$x_c = -f\frac{X_C}{Z_C} \quad y_c = -f\frac{Y_C}{Z_C}$$

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

$$\begin{bmatrix} T_C Z_C x_c \\ T_C Z_C y_c \\ T_C Z_C \end{bmatrix} = \begin{bmatrix} -f & 0 & 0 & 0 \\ 0 & -f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} T_C X_C \\ T_C Y_C \\ T_C Z_C \\ T_C \end{bmatrix} \quad (1)$$

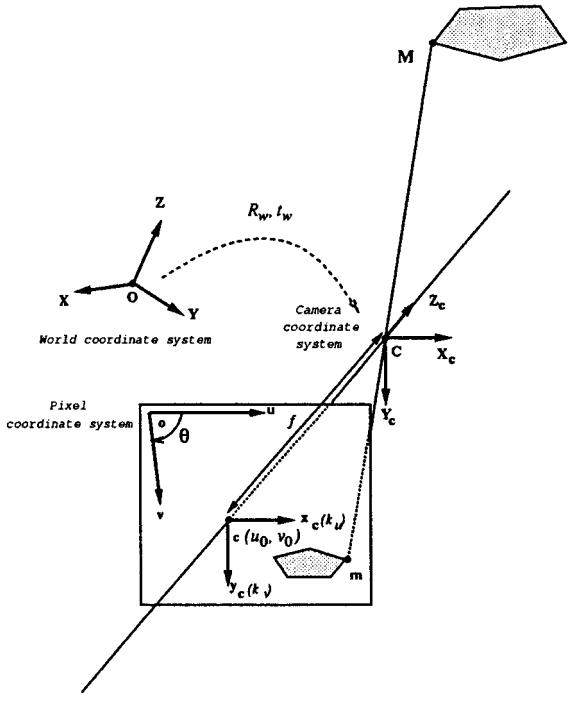

Figure 1. The general projective camera model.

In this equation $Z_C x_c$, $Z_C y_c$ and Z_C should be considered as the projective coordinates X_c, Y_c, Z_c of the pixel m and $T_C X_C$, $T_C Y_C$, $T_C Z_C$, T_C as the projective coordinates $\mathcal{X}_C$, $\mathcal{Y}_C$, $\mathcal{Z}_C$, $\mathcal{T}_C$ of the point M. We verify on this equation that the projective coordinates are defined up to a scale factor since multiplying them by an arbitrary nonzero factor does not change the Euclidean coordinates of either m or M. 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. In particular we have indicated in Fig. 1 another world coordinate system (O, X, Y, Z) and another retinal coordinate system (o, u, v). The coordinate system (O, X, Y, Z) is related to the coordinate system (C, X_C, Y_C, Z_C) by a rigid displacement described by the rotation matrix $\mathbf{R}$ and the translation vector $\mathbf{t}$. If we think of (O, x, y, z) as the laboratory coordinate system, the displacement describes the pose of the camera in the laboratory. The parameters describing the displacement are called the *extrinsic camera parameters*. The coordinate system (o, u, v) is related to the the coordinate system (c, x_c, y_c) by a change of scale of magnitude k_u and k_v along the u- and v-axes, respectively, followed by a translation $[u_0, v_0]^T$. In addition, the second coordinate axis is rotated from θ around o. The parameters relating the two retinal coordinate systems do not depend on the pose of the camera and are called the *camera intrinsic parameters*. The coordinate system (o, u, v) is the coordinate system that we use when we address the pixels in an image. It is usually centered at the upper left hand corner of the image which is usually not the point c, the pixels are usually not square and have aspect ratios depending on the actual size of the photosensitive cells of the camera as well as on the idiosyncrasies of the acquisition system. For most of the imaging situations which are commonly encountered, the retinal axes are orthogonal, and therefore the angle θ is zero. However, in order to be totally general this angle has to be considered for several reasons. First, the projection matrices, being 3×4 matrices defined up to a scale factor, depend on 11 free parameters. Since a displacement is described by 6 parameters (3 for the rotation, 3 for the translation), there are 5 intrinsic parameters, and therefore, in addition to the scale factors and the coordinates of the principal point, one additional intrinsic parameter is needed to ensure the existence of the decomposition of any projection matrix into extrinsic and intrinsic parameters. Second, from a practical standpoint, there are a couple of special imaging situations which cannot be properly described without considering θ. These situations include not only the case of pixel grids which have a non-orthogonal arrangement, but also the images taken with a bellows camera, or enlarged with a tilted easel (in both of these cases the optical axis may not be orthogonal to the image plane), or the case when pictures of pictures are considered. Even if the value of θ is known, if we include its recovery into the computation scheme, we obtain this way an easy quality check on the results.

This camera model is essentially linear and ignores nonlinear effects such as those caused by lens distortions. If such a correction is needed, it can be performed in a way compatible with the projective linear framework which is presented in this paper, as demonstrated by the work of Brand et al. (1993). Moreover, we believe that for several computer vision applications, such a correction is not even needed. The self-consistency of the results that we have obtained, as well as the quantitative reconstruction results show that the projective linear model is indeed appropriate, although the cameras which were used in the experiments were

off-the-shelf cameras normally priced, described in more details later. Most of the distortion problems can be avoided either with a reasonable choice of optics (care has to be taken using some extreme wide angles and some zoom lenses), or by using mostly points which lie in the center of the field of view. There are also a number of high-quality lenses, some of which quite cheap, for which the distortion is neglectible. In our own experiments with off-the-shelf cameras, as well as those conducted by Brand et al. (1993) it was found that the correction to be applied in order to obtain a projective linear model was subpixellic. Unless feature detectors reach the same precision, which is not always guaranteed even with state-of-the art algorithms, the additional correction is not very useful.

The fact that no nonlinear camera distortion is considered allows us to use the powerful tools of projective geometry. Projective geometry is emerging as an attractive framework for computer vision (Mundy and Zisserman, 1992). 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 (Coxeter, 1987; Garner, 1981; Semple and Kneebone, 1979), but also in the computer vision literature where it is presented in chapters of recent books (Faugeras, 1993; Kanatani, 1992; Mundy and Zisserman, 1992), and articles (Kanatani, 1991; Maybank and Faugeras, 1992).

Using Eq. (1) and the basic properties of changes of coordinate systems, we can express the relation between the image coordinates in the (o, u, v) coordinate system and the three-dimensional coordinates in the (O, x, y, z) coordinate system by the following equation

$$\begin{bmatrix} U \\ V \\ W \end{bmatrix} = \mathbf{A} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \mathbf{D} \begin{bmatrix} \mathcal{X} \\ \mathcal{Y} \\ \mathcal{Z} \\ \mathcal{T} \end{bmatrix} = \mathbf{P} \begin{bmatrix} \mathcal{X} \\ \mathcal{Y} \\ \mathcal{Z} \\ \mathcal{T} \end{bmatrix} \quad (2)$$

where U, V, and W are retinal projective coordinates, $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$, and $\mathcal{T}$ are projective world coordinates, $\mathbf{A}$ a 3×3 matrix describing the change of retinal coordinate system, and $\mathbf{D}$ is a 4×4 matrix describing the change of world coordinate system. The 3×4 matrix $\mathbf{P}$ is the perspective projection matrix, which relates 3-D world projective coordinates and 2-D retinal projective coordinates. Except for the points at infinity in the retina for which $W = 0$, the usual retinal coordinates u, v are related to the retinal projective coordinates by

$$u = \frac{U}{W} \quad v = \frac{V}{W}$$

The points at infinity in the retinal plane can be considered as the images of the 3-D points in the focal plane of the camera, i.e., the plane going through C and parallel to the retinal plane.

Similarly, except for the points at infinity in 3-D space for which $\mathcal{T} = 0$, the usual space coordinates X, Y, and Z are related to the projective world coordinates by

$$X = \frac{\mathcal{X}}{\mathcal{T}} \quad Y = \frac{\mathcal{Y}}{\mathcal{T}} \quad Z = \frac{\mathcal{Z}}{\mathcal{T}}.$$

The matrix $\mathbf{A}$ can be expressed as the following function of the intrinsic parameters and the focal length f

$$\mathbf{A} = \begin{bmatrix} -fk_u & fk_u \cot\theta & u_0 \\ 0 & -\frac{fk_v}{\sin\theta} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3)$$

Note that the focal length and minus sign appearing in (1) has been "transfered" to matrix $\mathbf{A}$ which depends on the products fk_u, fk_v which says that we cannot discriminate between a change of focal length and a change of units on the pixel axes. For this reason, we introduce the parameters $\alpha_u = -fk_u$ and $\alpha_v = -fk_v$. If $\theta = \pi/2$, Eq. (3) takes the simpler form:

$$\mathbf{A} = \begin{bmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (4)$$

Matrix $\mathbf{D}$ depends on 6 extrinsic parameters, three defining the rotation, three defining the translation, and has the form:

$$\mathbf{D} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}_3^T & 1 \end{bmatrix} \quad (5)$$

There is an interesting and important relationship between the camera intrinsic parameters and the absolute conic which is central to the problematic of this paper and which we study now. The absolute conic was used in (Faugeras and Maybank, 1990) to compute the number of solutions to the problem of estimating the motion of a camera from five point correspondences

in two views and in (Maybank and Faugeras, 1992) to study the problem of camera calibration. The absolute conic Ω lies in the plane at infinity of equation $\mathcal{T} = 0$ and its equation is

$$\mathcal{X}^2 + \mathcal{Y}^2 + \mathcal{Z}^2 = 0 \tag{6}$$

All points on that conic have complex coordinates. In fact, if we define $x = \frac{\mathcal{X}}{\mathcal{Z}}$ and $y = \frac{\mathcal{Y}}{\mathcal{Z}}$, the equation can be rewritten $x^2 + y^2 = -1$ which shows that it represents a circle of radius $i = \sqrt{-1}$. Even though this seems a little bit farfetched, this conic is closely related to the problem of camera calibration and motion estimation because it has the fundamental property of being invariant under rigid displacements, a fact already known to Cayley. The proof of this can be found in (Faugeras, 1993; Faugeras and Maybank, 1990). Let us examine the consequences of this invariance. Since the absolute conic is invariant under rigid displacements, its image by the camera, which is also a conic with only complex points, does not depend on the pose of the camera. Indeed, when we move the camera, the absolute conic does not change since it is invariant under rigid displacements and hence its image by the camera remains the same. Therefore, its equation in the retinal coordinate system (o, u, v) does not depend on the extrinsic parameters and depends only on the intrinsic parameters. By taking the identity displacement as extrinsic parameters, it is easily seen (Faugeras, 1993; Luong, 1992), that the matrix defining the equation of the image of the absolute conic in the retinal coordinate system (o, u, v) is:

$$\mathbf{B} = \mathbf{A}^{-T}\mathbf{A}^{-1} \tag{7}$$

One of the important ideas which has emerged from our previous work (Faugeras et al., 1992; Faugeras and Maybank, 1990; Maybank and Faugeras, 1992) and will also become apparent in this paper, is that the absolute conic can be used as a calibration pattern for the camera. This calibration pattern has the nice properties of always being present and of being free.

2.2. *The Epipolar Correspondence, the Fundamental Matrix and the Essential Matrix*

In the previous section, we have discussed the geometry of one camera. We are now going to introduce a second camera and study the new geometric properties of a set of two cameras. The main new geometric property is known in computer vision as the epipolar constraint and can readily be understood by looking at Fig. 2.

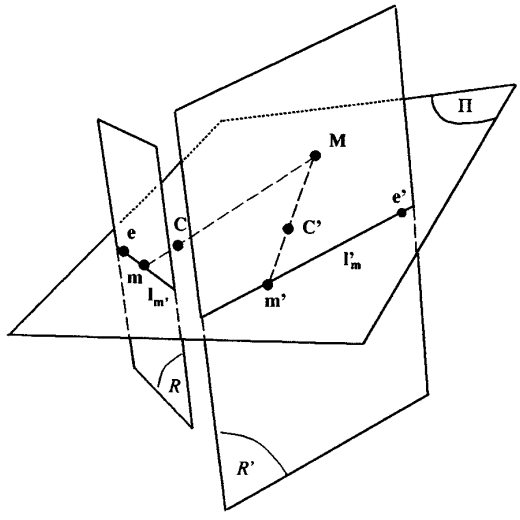

Figure 2. The epipolar geometry.

Let $\mathbf{C}$ (resp. $\mathbf{C}'$) be the optical center of the first camera (resp. the second). The line $\langle \mathbf{C}, \mathbf{C}' \rangle$ projects to a point $\mathbf{e}$ (resp. $\mathbf{e}'$) in the first retinal plane $\mathcal{R}$ (resp. in the second retinal plane $\mathcal{R}'$). The points $\mathbf{e}$, $\mathbf{e}'$ are the epipoles. The lines through $\mathbf{e}$ in the first image and the lines through $\mathbf{e}'$ in the second image are the epipolar lines. The epipolar constraint is well-known in stereovision: for each point $\mathbf{m}$ in the first retina, its corresponding point $\mathbf{m}'$ lies on its epipolar line $\mathbf{l}'_m$.

A 2D point, as well as a 2D line, is represented in projective geometry by a vector of three coordinates. Two proportional vectors represent the same point (or line). The point $\mathbf{m} = [m_1, m_2, m_3]^T$ belongs to the line $\mathbf{l} = [l_1, l_2, l_3]^T$ if, and only if $\mathbf{l}^T\mathbf{m} = 0$. The key observation is that the relationship between the retinal coordinates of a point $\mathbf{m}$ and its corresponding epipolar line $\mathbf{l}'_m$ is projective linear. The fundamental matrix describes this correspondence:

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

The epipolar constraint has then a very simple expression: since the point $\mathbf{m}'$ corresponding to $\mathbf{m}$ belongs to the line $\mathbf{l}'_m$ by definition, it follows that

$$\mathbf{m}'^T\mathbf{F}\mathbf{m} = 0 \tag{8}$$

The epipoles e and e' are special points which verify the following relations:

$$\mathbf{F}\mathbf{e} = \mathbf{F}^T\mathbf{e}' = \mathbf{0} \tag{9}$$

They imply that the rank of $\mathbf{F}$ is less than equal to 2, and in general it is equal to 2. Since the matrix is defined up to a scale factor, it depends upon seven independent parameters.

Equation (8) is the analog in the uncalibrated case of the so-called Longuet-Higgins equation (Longuet-Higgins, 1981). Indeed, in the case of calibrated cameras, the 2D projective coordinates of a point $\mathbf{m}$ give the 3-D direction of the optical ray $\mathbf{Cm}$ (see Fig. 2), which is of course not the case with retinal (uncalibrated) coordinates. If the motion between the two positions of the camera is given by the rotation matrix $\mathbf{R}$ and the translation matrix $\mathbf{t}$, and if $\mathbf{m}$ and $\mathbf{m}'$ are corresponding points, then the coplanarity constraint relating $\mathbf{Cm}'$, $\mathbf{t}$, and $\mathbf{Cm}$ is written as:

$$\mathbf{m}' \cdot (\mathbf{t} \times \mathbf{Rm}) \equiv \mathbf{m}'^T \mathbf{Em} = 0 \tag{10}$$

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, i.e., the translation vector is defined up to a scale factor.

It can be seen that the two Eqs. (8) and (10) are equivalent, and that we have the relation:

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

Unlike the essential matrix, which is characterized by the two constraints found by Huang and Faugeras (1989) which are the nullity of the determinant and the equality of the two non-zero singular values, the only constraint on the fundamental matrix is that it is of rank two.

2.3. *The Rigidity Constraint, Kruppa Equations and the Intrinsic Parameters*

We now provide a direct link between the fundamental matrix (a projective quantity), and the intrinsic parameters (Euclidean quantities). Several formulations are presented and proved to be equivalent in spite of the intriguing discrepancy in the number of constraints that each of them generates.

Algebraic Formulations of the Rigidity Constraints Using the Essential Matrix. In the case of two different cameras, the transformation between the two retinal coordinate systems is a general linear projective transformation of $\mathcal{P}^3$, depending on 15 parameters. This transformation can be decomposed in two (possibly similar) changes of retinal coordinates, and one rigid displacement. This decomposition is very far from being unique. However, not all the choices of the intrinsic parameters are possible. The constraints on the intrinsic parameters are obtained by expressing the rigidity of this underlying displacement, the fact that for any fundamental matrix $\mathbf{F}$, one can find intrinsic parameters matrices $\mathbf{A}$ and $\mathbf{A}'$, such that $\mathbf{A}'^T\mathbf{FA}$ is an essential matrix. We have seen that only the seven parameters of the fundamental matrix are available to describe the geometric relationship between two views. The five parameters of the essential matrix are needed to describe the rigid underlying displacement between the associated normalized coordinate systems, thus we can see that at most two independent constraints are available for the determination of intrinsic parameters from the fundamental matrix.

A first set of approaches to express the rigidity constraint involve the essential matrix:

$$\mathbf{E} = \mathbf{A}'^T\mathbf{FA} \tag{11}$$

The rigidity of the motion yielding the fundamental matrix $\mathbf{F}$ with intrinsic parameters $\mathbf{A}$ and $\mathbf{A}'$ is equivalent to the Huang and Faugeras conditions expressing the fact that $\mathbf{E}$, defined by (11) is an essential matrix:

$$\begin{aligned} &\det(\mathbf{E}) = 0 \\ &f(\mathbf{E}) \;=\; \frac{1}{2}\mathrm{trace}^2(\mathbf{EE}^T) - \mathrm{trace}(\mathbf{EE}^T)^2 = 0 \end{aligned} \tag{12}$$

As we have $\det(\mathbf{F}) = 0$, the first condition is automatically satisfied, and does not yield any valuable constraint in our framework, thus we are left with only one polynomial constraint, the second condition.

A second expression of the rigidity constraints has been presented by Trivedi (1988). If $\mathbf{E}$ is an essential matrix, the symmetric matrix $\mathbf{S} = \mathbf{EE}^T$, which *a priori* has six independent entries, depends only on the three components of $\mathbf{t}$:

$$\mathbf{EE}^T = -[\mathbf{t}]_\times{}^2 = \begin{bmatrix} t_2^2 + t_3^2 & -t_1t_2 & -t_1t_3 \\ -t_2t_1 & t_3^2 + t_1^2 & -t_2t_3 \\ -t_3t_1 & -t_3t_2 & t_1^2 + t_2^2 \end{bmatrix} \tag{13}$$

The matrix $\mathbf{S} = \mathbf{EE}^T$ has thus a special structure in which the three diagonal and the three off-diagonal entries are related by the three relations designated by (T_{ij}), $1 \le i < j \le 3$:

$$4S_{ij} - (\mathrm{trace}(\mathbf{S}) - 2S_{ii})(\mathrm{trace}(\mathbf{S}) - 2S_{jj}) = 0 \; (T_{ij})$$

Trivedi has shown that in the case he considered, where the only intrinsic parameters were the coordinates of the principal point, his three polynomial constraints reduce in fact to a tautology and two independent polynomial constraints, provided that $\det(\mathbf{E}) = 0$. An examination of his proof shows that this fact is true in the case of a general intrinsic parameters model too. Thus we are left with two polynomial constraints, in addition to the nullity of the determinant.

We show in Appendix A that in spite of the apparent discrepancy in the number of equations, these approaches to express the rigidity are equivalent. However, the two independent Trivedi equations which are equivalent to the second Huang and Faugeras condition are not simpler than this one, contradicting what would be expected. They all yield algebraic constraints which are polynomials of degree 8 in the coefficients of $\mathbf{A}$ and $\mathbf{A}'$ (the intrinsic parameters) and thus are not suitable for practical computation, or even theoretical study. It is why we are going to consider a *geometrical* interpretation of the rigidity constraint which yields low-order polynomial constraints.

The Kruppa Equations: A Geometric Interpretation of the Rigidity Constraint. The Kruppa equations (Kruppa, 1913) are obtained from a geometric interpretation of the rigidity constraints. They were first introduced in the field of computer vision by Faugeras and Maybank for the study of motion (Faugeras and Maybank, 1990), and then to develop a theory of self-calibration (Maybank and Faugeras, 1992). In this exposition, we will return to the original formulation, which doesn't assume that the two cameras are identical, which is useful to prove the equivalence of the Kruppa equations and the Huang and Faugeras constraint.

Let consider an epipolar plane Π, which is tangent to Ω. Then the epipolar line l is tangent to ω, projection of Ω into the first image, and the epipolar line l' is tangent to the projection ω' of Ω into the second image. It follows that the two tangents to ω from the epipole $\mathbf{e}$ correspond under the epipolar transformation to the two tangents to ω' from the epipole $\mathbf{e}'$, as illustrated by Fig. 3.

If $\mathbf{B}$ is the matrix of ω, image of the absolute conic in the first camera, then the matrix of the dual conic of ω (formed by the tangents to ω) is the dual matrix (matrix of cofactors) of $\mathbf{B}$: $\mathbf{K} = \mathbf{B}^*$. Since $\mathbf{B}^*$ is proportional to the inverse of $\mathbf{B}$, and since we are dealing with matrices defined up to a scale factor, we can take, using (7):

$$\mathbf{K} = \mathbf{A}\mathbf{A}^T \tag{14}$$

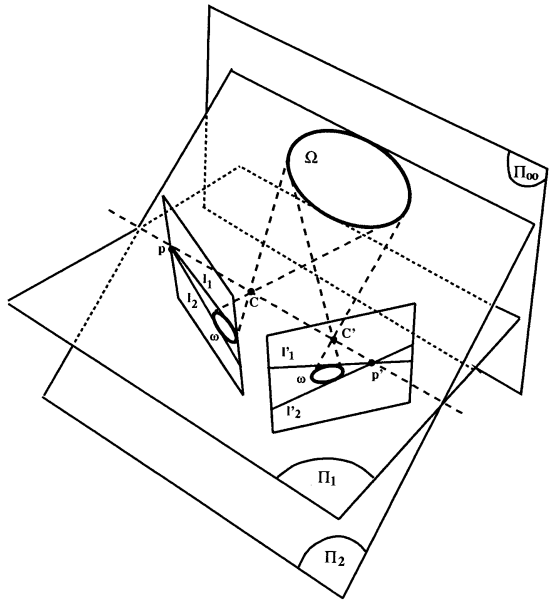

Figure 3. The absolute conic and epipolar transformation.

This makes explicit the fact that the matrix $\mathbf{A}$ can be obtained uniquely from the Cholesky decomposition (Golub and Van Loan, 1989) of the symmetric matrix $\mathbf{K}$ when this matrix is positive definite, which is always the case since ω has only complex points. A remark which will be useful later is that when the pixel grid is orthogonal ($\theta = \frac{\pi}{2}$), then:

$$K_{13}K_{23} - K_{33}K_{11} = 0 \tag{15}$$

The epipolar line $l = \langle \mathbf{e}, \mathbf{y} \rangle$ is tangent to ω iff:

$$(\mathbf{e} \times \mathbf{y})^T \mathbf{K} (\mathbf{e} \times \mathbf{y}) = 0 \tag{16}$$

The epipolar line corresponding to the point $\mathbf{y}$ is $\mathbf{F}\mathbf{y}$ and is tangent to ω' iff:

$$\mathbf{y}^T \mathbf{F}^T \mathbf{K}' \mathbf{F} \mathbf{y} = 0 \tag{17}$$

Writing that (16) and (17) are equivalent yield the so-called Kruppa equations.

One possible way to do so is to take $\mathbf{y} = (1, \tau, 0)^T$ in the case where the epipoles are at finite distance. The relations (16) (resp. (17)) take the form $P_1(\tau) = k_0 + k_1\tau + k_2\tau^2 = 0$ (resp. $P_2(\tau) = k'_0 + k'_1\tau + k'_2\tau^2 = 0$), and the Kruppa equations can be written as three proportionality conditions between these two

polynomials, of which only two are independent:

$$\begin{aligned} k_2k_1' - k_2'k_1 &= 0 \\ k_0k_1' - k_0'k_1 &= 0 \\ k_0k_2' - k_0'k_2 &= 0 \end{aligned} \tag{18}$$

We have shown that the Kruppa equations are equivalent to the Huang and Faugeras constraint expressed using the fundamental matrix and the intrinsic parameters. As seen previously, the null determinant constraint is readily satisfied, thus we have only to show that the set of Kruppa Eqs. (18) is equivalent to the second constraint, which is done in Appendix B. This is something which is a priori not intuitive, since there are *two* Kruppa equations.

The nice thing with the Kruppa equations is that since the coefficients k_i (resp. k_i') depend linearly on the entries of $\mathbf{K}$ (resp. $\mathbf{K}'$), they are only of degree *two* in these entries, thus providing a much simpler expression of the rigidity constraint than the one obtained by the purely algebraic methods described at the beginning of this section.

3. Using the Kruppa Equations to Compute the Intrinsic Parameters

In this section, we examine how to use in practice the two Kruppa equations relating fundamental matrices and intrinsic parameters which have been derived in the previous section. We first examine how many movements or images are necessary, as well as the cases of degeneracy. Two types of numerical methods to solve the equations are then considered, and simulations are presented to assess the strengths and weaknesses of each method.

3.1. Using Three Displacements of a Moving Camera

A Moving Camera. In an earlier work, Trivedi (1988) has considered the problem of computing only the coordinates of the principal point of each camera, that is to solve the self-calibration problem for the restricted model of intrinsic parameters:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & u_0 \\ 0 & 1 & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad \mathbf{A}' = \begin{bmatrix} 1 & 0 & u_0' \\ 0 & 1 & v_0' \\ 0 & 0 & 1 \end{bmatrix}$$

using the *three* equations (T_{ij}) mentioned previously. The initial idea was that if there were three such independent equations, then it would have been possible to find a solution as soon as the number of cameras is superior or equal to three. But Trivedi pointed out that the three equations reduce to two independent equations, and a tautology, and thus that there are not enough constraints for the problem to be solved.

Recently, Hartley (1992) has brought a partial solution using a simplified camera model, where the only unknown is the focal distance, thus taking as a model for the intrinsic parameters:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & k \end{bmatrix} \quad \mathbf{A}' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & k' \end{bmatrix}$$

He exhibits an algorithm to factor the fundamental matrix $\mathbf{F}$ as $\mathbf{A}'^T\mathbf{E}\mathbf{A}^{-1}$, which under his assumption depends also on seven parameters, the two different focal lengths and the five motion parameters.

If we do not make an additional assumption, it is not possible to use a more general model for the intrinsic parameters, since by adding views, we add a number of unknowns that is a least equal to the number of additional equations. The idea behind our method is to use constraints which arise from the observation of a static scene by a *single*[1] moving camera. In this case the intrinsic parameters remain constant: $\mathbf{A} = \mathbf{A}'$, $\mathbf{K} = \mathbf{K}'$, thus we can cumulate constraints over different displacements, and obtain a sufficient number of equations for the resolution.

How Many Displacements or Images are Necessary? Each displacement yields two independent algebraic equations. In the case of a moving camera, we have only five coefficients of $\mathbf{K}$ to estimate, since $\mathbf{K}$ is a symmetric matrix defined up to a scale factor. In the general case, three displacements are necessary. In the case of the simplified model with four intrinsic parameters, two displacements are sufficient, since we have the additional constraint (15). Note that in the most general case, the displacements need not be in a sequence, which means that we can consider for example the displacements 1-2, 3-4, 5-6 between 6 different images. However, the displacements are not always independent, when the images which are considered form a sequence. The algorithm based on the Kruppa equations ignores the additional constraints which arises then.

Between three images, there are three displacements, 1-2, 2-3, et 1-3. One could worry about the fact that

since the third displacement in this case $\mathbf{D}_3 = \mathbf{D}_2\mathbf{D}_1$ is a composition of the two first displacements $\mathbf{D}_1$ and $\mathbf{D}_2$, the 1-3 equations would be dependent on the 1-2 and 2-3 equations, thus resulting in an under-constrained system. One way to see that it is not the case is to count unknowns. Two fundamental matrices depend only upon 14 parameters. On the other hand, the Euclidean information which is recovered by self-calibration, consists of the three displacements 1-2, 2-3, and 1-3 up to a common scale factor, and the 5 intrinsic parameters. The displacements depend on 11 parameters: 3 for each of the rotations $\mathbf{R}_{12}$ and $\mathbf{R}_{23}$, 2 for each of the directions of translations $\mathbf{t}_{12}$ and $\mathbf{t}_{23}$, one for the ratio of the norms of the translations. The total is 16 parameters, thus the information is not entirely contained in the first two fundamental matrices. These two missing parameters are actually recovered thanks to the two additional Kruppa equations provided by the third fundamental matrix. We also give in Appendix C a simple numerical example to show that in the general case the equations are independent.

Degenerate Cases. Not all combinations of displacements will work. For instance, if two of the displacements are identical, obviously they will yield only two independent constraint.

Also, in the case of a displacement for which the translation vector is null $\mathbf{t} = (0, 0, 0)^T$, that is if the displacement is a pure rotation whose axis goes contains the optical center of the camera, as the two optical centers are identical, there is no epipolar constraint, and thus the rigidity constraint cannot be expressed by means of the Kruppa equations. However, self-calibration is even easier in this case, since instead of two quadratic constraints on the entries of $\mathbf{K}$, one obtains five linear constraints on these entries, as shown in (Hartley, 1994b; Luong and Viéville, 1993).

In the case where the displacement is a pure translation (the rotation is the identity $\mathbf{R} = \mathbf{I}_3$), it can be seen from (10) and (11) that the fundamental matrix is antisymmetric. From (9), we conclude that

$$\forall \mathbf{y},\ \mathbf{F}\mathbf{y} = \mathbf{e} \times \mathbf{y}$$

therefore the Eqs. (16) and (17) are equivalent, and thus the Kruppa equations reduce to tautologies. A geometric interpretation is that since the two tangents to ω in the first image *are* the tangents to ω in the second image, no further constraint is put on ω by considering a second image.

3.2. A Semi-Analytic Method

Principle. Three displacements yield six equations in the entries of the matrix $\mathbf{K}$. The equations are homogeneous, so the solution is determined only up to a scale factor. In effect there are five unknowns. Trying to solve the over-determined problem with numerical methods usually fails, so five equations are picked from the six and solved first. As the equations are each of degree two, the number of solutions in the general case is $32 = 2^5$. The remaining equation could just be used to discard the spurious solutions, but we have prefer red to exploit the redundancy of information to obtain a more robust algorithm, as well as a gross estimate of the variance of the solutions.

Solving the Polynomial System by Continuation. A problem is that solving a polynomial system by providing an initial guess and using an iterative numerical method will not generally find all the solutions: many of the starting points will yield trajectories that do not converge and many other trajectories will converge to the same solution. However it is not acceptable to miss solutions, since there is only one correct one amongst the 32. Recently developed methods in numerical continuation can reliably compute all solutions to polynomial systems. These methods have been improved over a decade to provide reliable solutions to kinematics problems. The details of these improvements are omitted. The interested reader is referred for instance to (Wampler et al., 1988) for a detailed tutorial presentation. The solution of a system of nonlinear equations by numerical continuation is suggested by the idea that small changes in the parameters of the system usually produce small changes in the solutions. Suppose the solutions to problem A (the start system) are known and solutions to problem B (the target system) are required. Solutions to the problem are tracked as the parameters of the system are slowly changed from those of A to those of B. Although for a general nonlinear system numerous difficulties can arise, such as divergence or bifurcation of a solution path, for a polynomial system all such difficulties can be avoided. Using an implementation provided by Jean Ponce and colleagues fairly precise solutions can be obtained. The major drawback of this method is that it is expensive in terms of CPU time. The method is a naturally parallel algorithm, because each continuation path can be tracked on a separate processor. Running it on a network of 7 Sun-4 workstations takes approximatively half a minute to solve one system of equations.

Continuation-based computation of the intrinsic parameters with three displacements

- generate six independent Kruppa equations from the three fundamental matrices.
- for each of the six system of five equations $\mathcal{E}_i$:
  - — solve $\mathcal{E}_i$ by the continuation method to obtain the Kruppa matrices $\mathbf{K}_i^m$ with $1 \leq m \leq 32$
  - — keep only the matrices $\mathbf{K}_i^m$ which have real entries and are positive
- for each list i of solutions, $(i = 1 \cdots 6)$
  - — for each list j of solutions, $(j = i + 1 \cdots 6)$
    - $*$ find the solution $\mathbf{K}_i^m$ in the list i and the solution $\mathbf{K}_j^n$ in the list j which minimize the distance[2]:

$$d(\mathbf{u}, \mathbf{v}) = \sum_{l=1}^{5} \frac{|u_l - v_l|}{\max(|u_l|, |v_l|)}$$

      where $\mathbf{u}$ and $\mathbf{v}$ are two 5-dimensional vectors representing $\mathbf{K}_i^m$ and $\mathbf{K}_j^n$, and u_l and v_l are their components,
    - $*$ increment the counters corresponding to $\mathbf{K}_i^m$ and $\mathbf{K}_j^n$
- choose in each list the solution which has obtained the highest counter score,
- compute the intrinsic parameters
- compute the final solution and an estimate of the covariance by an averaging operator[3].

Two Examples. We present in Tables 1 and 2 two typical examples. They differ by the choice of the three movements. It can be seen that results can differ significantly from one configuration to another. The second configuration yields results which are less satisfactory because one of the three displacements has a rather small rotational component.

The experimental procedure consisted in choosing three displacements, generating point correspondences by projecting in the two retinas a set of random 3D points and adding Gaussian image noise, computing the fundamental matrix with a non-linear method (Luong et al., 1993) from these point correspondences, and then use the continuation algorithm to solve the Kruppa equations obtained from the fundamental matrices. To match real conditions, the size of the images was taken to be 512×512 pixels, the field of view of the camera was 43°, and the 3D points were scattered in a cube of size 10 meters centered around the initial position of the camera.

We have verified the impact of the orthogonality constraint (15) in two different cases. In the first one, where we have only two sets of correspondences, we just solve the system of equations which are the four Kruppa equations and the constraint (15), which is also quadratic. In the second one, where we have three sets of correspondences, we could use a very strong redundancy of equations since there are now $C_6^4 = 15$ systems which could be built by picking four Kruppa equations plus the constraint (15). However, for the sake of comparison, we have used only six of these equations.

We have tested:

- 2 displacements with the orthogonality constraint (the results are displayed with the three possible combinations of displacements, the dash indicate that no solution compatible with the constraints of positivity of $\mathbf{K}$ was found),
- 3 displacements with the orthogonality constraint,
- 3 displacements without the orthogonality constraint.

Numbers in brackets are estimates of the uncertainty of the results.

The big advantage of the method is that no initialization is needed. If the points are measured with a good precision, the results can be sufficiently precise. Another advantage is that it is easy to assess the success or failure of the algorithm. However there are several drawbacks:

Table 1. Results obtained with the continuation method, configuration 1.

Noise (pixels)	Method		Estimated parameters				
	Orth.	Displ.	α_u	α_v	u_0	v_0	$\theta - \frac{\pi}{2}$
0			640.125	943.695	246.096	255.648	0
0.1	Y	1,2	642.32	947.37	245.82	253.94	
	Y	2,3	639.44	944.36	246.04	258.59	
	Y	1,3	641.62	945.73	248.97	255.56	
	Y	1,2,3	641.69 [2.0]	947.49 [3.7]	247.03 [1.2]	256.55 [1.7]	
	N	1,2,3	644.40 [2.3]	952.29 [3.8]	237.45 [4.2]	254.61 [1.9]	
0.5	Y	1,2	651.39	962.47	244.73	246.56	
	Y	2,3	636.54	946.72	245.82	270.46	
	Y	1,3	647.41	953.67	260.91	255.11	
	Y	1,2,3	648.39 [11.1]	963.69 [20.3]	250.84 [6.7]	260.26 [9.1]	
	N	1,2,3	664.19 [11.2]	996.03 [20.6]	190.91 [23.8]	248.71 [9.1]	4.10^{-2} [2.10^{-2}]
1.0	Y	1,2	—	—	—	—	—
	Y	2,3	632.49	948.90	245.50	285.45	
	Y	1,3	74.85	455.95	733.93	434.07	
	Y	1,2,3	658.00 [24.8]	986.63 [45.7]	255.61 [14.3]	265.09 [19.7]	
	N	1,2,3	681.66 [25.7]	1109.05 [75.6]	31.10 [139.9]	231.99 [20.5]	0.13 [0.08]
1.5	Y	1,2	676.05	1002.37	241.89	223.28	
	Y	2,3	627.92	950.11	245.16	300.56	
	Y	1,3	659.79	971.19	293.82	252.73	
	Y	1,2,3	669.62 [42.6]	1013.85 [79.2]	260.16 [23.3]	270.23 [32.3]	
	N	1,2,3	633.02 [73.0]	1223.62 [104.5]	190.46 [231.1]	205.49 [43.9]	0.27 [0.2]

- the method is suitable only for the case of the minimum number of displacement, as it is difficult to use all the constraints provided by a long sequence without increasing considerably the amount of computations,
- it is difficult to take into account uncertainty for the input (fundamental matrices) as well as for the output (camera parameters),
- the computational cost of solving the polynomial system is relatively high,
- it is not possible to express the constraints of positivity of **K** at the resolution level, since continuations work in the complex plane. Thus with noisy data, it can happen that no acceptable solution can be found,
- it is not very easy to use some a priori knowledge that one might have about the intrinsic parameters.

All these drawbacks come from the use of the continuation method and can be overcome using an optimization formulation.

3.3. Optimization Formulation Taking into Account Multiple Views

In this approach, we no longer make use of the simple polynomial structure of the Kruppa equations, but rather consider them as measurement equations relating directly the fundamental matrices to intrinsic parameters, obtained by substituting the values of the entries of **K** obtained from (14) into (18). We can then solve them either by a batch non-linear least-squares minimization technique, or by an extended Kalman filtering approach. In both cases, the uncertainty of the measurements (**F**) can be taken into account.

Global Minimization. The choice of the error function to be minimized is very important. We have noticed two things. First, using the three Kruppa Eqs. (18) even though they are not independent provides additional constraints and improve the results. Second, minimizing directly the value of the residual

Table 2. Results obtained with the continuation method, configuration 2.

Noise (pixels)	Method		Estimated parameters				
	Orth.	Displ.	α_u	α_v	u_0	v_0	$\theta - \frac{\pi}{2}$
0			640.125	943.695	246.096	255.648	0
0.1	Y	1,2	647.56	955.50	245.32	250.58	
	Y	2,3	124.317	947.934	230.705	252.053	
	Y	1,3	639.303	943.591	246.083	257.594	
	Y	1,2,3	640.83 [2.7]	947.59 [2.7]	237.90 [10.2]	252.88 [3.1]	
	N	1,2,3	636.32 [15.8]	942.45 [5.0]	241.87 [5.9]	251.61 [2.8]	0.018 [0.02]
0.5	Y	1,2	—	—	—	—	—
	Y	2,3	—	—	—	—	—
	Y	1,3	635.76	942.88	246.03	265.70	
	Y	1,2,3	654.01 [24.1]	976.83 [22.4]	214.28 [47.0]	232.66 [20.7]	
	N	1,2,3	623.63 [78.8]	934.15 [31.4]	240.84 [2.1]	237.95 [14.7]	0.089 [0.09]
1.0	Y	1,2	744.34	1110.38	235.28	187.89	
	Y	2,3	—	—	—	—	—
	Y	1,3	630.86	941.71	245.96	277.04	
	Y	1,2,3	505.94 [248.7]	779.03 [389.9]	179.30 [94.4]	407.68 [317.3]	
	N	1,2,3	628.20 [130.4]	936.94 [68.6]	208.05	217.74 [27.9]	0.15 [0.1]
1.5	Y	1,2	2462.05	3943.05	27.53	−558.13	
	Y	2,3	342.86	875.35	219.38	246.91	
	Y	1,3	604.46	885.15	249.27	260.23	
	Y	1,2,3	688.43 [163.9]	1048.77 [254.3]	161.48 [75.0]	207.08 [38.7]	
	N	1,2,3	1190.91 [1164.0]	1803.80 [1867.3]	109.39 [149.1]	−109.65 [661.5]	0.13 [0.1]

of expressions (18) do not work well. The reason is the well known fact that minimizing an error function $\sum(\frac{a_i}{b_i} - \frac{a'_i}{b'_i})^2$ is quite different from minimizing $\sum(a_i b'_i - a'_i b_i)^2$ because the later is weighted by the variable quantity $b_i b'_i$. In our case, since we are interested in expressing the proportionality of the polynomials P_1 and P_2 (defined just before (18)) we thus minimize the following error function:

$$\min_{\alpha_u,\alpha_v,u_0,v_0,\theta} \sum_{\text{displacements}} \left(\frac{k_0}{k'_0} - \frac{k_1}{k'_1}\right)^2 + \left(\frac{k_1}{k'_2} - \frac{k_1}{k'_2}\right)^2 + \left(\frac{k_0}{k'_2} - \frac{k_0}{k'_2}\right)^2 \qquad (19)$$

where the coefficients k_i and k'_i are defined just before (18).

We have compared this method with the continuation method, using a statistical approach involving 100 triples[4] of displacements. To obtain an idea of the precision and convergence properties, we have started the minimization with different initial values: (1) the exact values, (2) the values given by the continuation method, (3) the arbitrary values $\alpha_u = 800$, $\alpha_v = 800$, $u_0 = 255$, $v_0 = 255$, $\theta = \frac{\pi}{2}$, corresponding to the relatively standard situation of an orthogonal pixel grid, no principal point shift, and reasonable values for the scale factors (with no knowledge of aspect ratio). The Table 3 shows the mean relative error for each parameter, obtained at two different noise levels: 0.2 is approximately the subpixel precision of the model-based feature detectors, whereas 1.0 pixel is the typical precision of some operator-based feature detectors.

We can conclude from these results that:

- the precision on the scale factors α_u and α_v is better than the one on the principal point coordinates u_0 and v_0,

Table 3. Statistical comparison of the continuation method and of optimization methods with 3 displacements. The minimization is initialized with: (1) exact values, (2) the continuation method, (3) an arbitrary value. The table shows the percent average error.

Noise (pixels)	Method	Failure	Parameters				
			α_u	α_v	u_0	v_0	$\theta - \frac{\pi}{2}$
0.2	Continuation	2	5	6	3	3	3
	Mini (1)		2	3	7	9	3
	Mini (2)		5	5	10	12	5
	Mini (3)		6	6	13	15	5
1.0	Continuation	7	12	14	26	32	8
	Mini (1)		11	14	26	32	9
	Mini (2)		13	16	27	37	13
	Mini (3)		17	18	28	36	13

- the results are quite sensitive to the choice of the initialization point,
- the precision of the iterative method is roughly comparable with the precision of the continuation method,

Since the number of equations and parameters is relatively small, the method is computationally efficient. Its main disadvantage is the need for a good starting point, but it could be obtained by the continuation method.

Recursive Filtering. If we have a long sequence, it may be interesting to use the Iterated Extended Kalman Filter[5], with the following data:

vector of state parameters

$$\mathbf{a} = (\alpha_u, \alpha_v, u_0, v_0)^T$$

vector of measurements

$$\mathbf{x} = (F_{11}, F_{12}, F_{13}, F_{21}, F_{22}, F_{23}, F_{31}, F_{32}, F_{33})^T$$

measurement equations

$$\mathbf{f}(\mathbf{x}, \mathbf{a}) = \mathbf{0},\ f_1 \text{ and } f_2 \text{ are two Kruppa Eqs. (18)}$$

The orthogonality correction factor has been dropped to reduce non-linearities in the model, and we have only used two Kruppa equations to ensure that the measurement equations are independent. The uncertainty on the fundamental matrices is needed. It is obtained using the method described in (Luong and Faugeras, 1994, 1995).

Statistical results have been conducted to see the effect of the increase of the number of displacements and to compare the Kalman method to the batch minimization approach[6]. In Table 4 the Kalman filtering has been initialized with the parameters estimated from the minimization technique using the first three displacements. The fact that the average error remains approximatively the same for the parameters α_u and α_v is due to convergence to false local minima induced by inexact starting points, and the fact that in the Kalman filter approach, the full information provided by all the displacements is not available, due to the recursive nature of the approach.

Thus, statistically, the global minimization gives better results, a finding consistent with those of (Kumar et al., 1989) and (Weng et al., 1989). However, if the starting point is precise, as in Table 5, where it is found by the minimization method using a larger number of displacements, it can be seen that the results are slightly better, which may be due to the fact that uncertainty is

Table 4. Statistical comparison of minimization and Kalman filtering. Initialization is done with minimization on 3 displacements. The table show the percent average error.

Noise (pixels)	nb (displ)	α_u		α_v		u_0		v_0	
		Mini	Kalman	Mini	Kalman	Mini	Kalman	Mini	Kalman
0.2	3	14	13	14	13	30	27	30	28
	5	9	13	8	14	22	26	25	27
	10	6	13	7	13	19	23	24	24
	15	4	13	7	12	20	22	20	21
1.0	3	38	35	39	36	52	50	57	55
	5	33	33	31	36	47	49	49	52
	10	28	35	29	37	44	46	51	48
	15	30	36	27	37	43	43	49	47

Table 5. Statistical comparison of minimization and Kalman filtering. Initialization done with the minimization on 15 displacements. The table show the percent average error, and the percent of tries for which the final error is above 5%.

Noise (pixels)		α_u		α_v		u_0		v_0	
		Mini	Kalman	Mini	Kalman	Mini	Kalman	Mini	Kalman
0.2	err	4	4	7	6	18	14	19	13
	% err > 5%	22	13	28	15	48	29	58	40
1.0	err	23	24	22	27	41	35	48	41
	% err > 5%	68	49	64	58	78	59	78	72

taken into account. In this table, we have mentioned not only the average relative error, but also the percentage of cases for which the final error was superior to 5%, which shows that if the Kalman filter does not fall into a false minimum, it improves the results significantly.

3.4. An Evaluation of the Methods Based on the Kruppa Equations

From the numerous simulations that we performed (some of which were described in this section), it appears that all the methods give results which are comparable, in the sense that none of them gives clearly superior results in all situations. In any case, the main limitation of the method comes from the necessity to get precise localization of the points in order to compute precise fundamental matrices. A subpixel accuracy of about 0.2 to 0.5 pixel is necessary in order to get acceptable results. It means that the most precise feature detectors need to be used. Some types of displacements will not work well, specifically those leading to nearly degenerate cases for Kruppa equations, mentioned in this section, and those leading to unstable computation of the fundamental matrix, which are studied in (Luong, 1992; Luong and Faugeras, 1995).

Another limitation might be that the method does not give an accurate estimation for the position of the principal point, and the angle of the retinal axes. The later is of no importance, since in practice it is very well controlled and very close to $\frac{\pi}{2}$. Thus this information can be used, either to restrict the model, or to discard false solutions. We will see in the next section that the former is also of little importance, in the sense that it does not affect a lot the subsequent stage of the calibration, the estimation of 3D motion. In fact, we will see that even with imprecise values of the camera parameters, fairly acceptable motion parameters can be recovered, and that furthermore, during this process of recovering the motion parameters, the estimation of intrinsic parameters can be refined.

To summarize the findings of this section, we recommend to use the continuation method when there is only a mininal number of displacements available, or when there is no initial guess available for the intrinsic parameters. When more displacements are available, and provided that there is a reasonable initial guess, we recommend to use the optimization approaches. The choice between the batch non-linear minimization algorithm and the Kalman filter algorithm depends on several factors. The advantages of the Kalman filter algorithm are that it is incremental, very fast, and gives the most precise results if the initialization point is very precise. On the other hand, it is considerably more sensitive to the choice of the initialization point than the batch minimization algorithm.

4. Taking into Account the Motion of the Camera

Our first goal in this section is to compute the three-dimensional motion from pairs of images, supposing that we have obtained the intrinsic parameters **A**. We show experimentally that this computation can be done quite robustly even with imprecise camera parameters, provided that the appropriate algorithm is chosen. The interesting remark is that the algorithm which is the most robust to image noise is not the one which is the most robust to imprecision of intrinsic parameters. The comparison of the motion determination algorithms is the basis for a new self-calibration algorithm. By combining the computation of motion with the computation of the intrinsic parameters we obtain another iterative approach to self-calibration, which we compare to the Kruppa approach.

4.1. Computing the Motion after Calibrating

The motion determination problem from point correspondences is very classical. See (Faugeras et al., 1987; Horn, 1990; Spetsakis and Aloimonos, 1988; Weng

et al., 1989) for solutions similar to ours. There are two different solutions, both based on the computation of the fundamental matrix.

A Direct Factorization. We have seen that during the course of intrinsic parameters estimation, we had to compute the fundamental matrix **F**, from which the essential matrix is immediately obtained:

$$\mathbf{E} = \mathbf{A}^T \mathbf{F} \mathbf{A} \tag{20}$$

The problem of finding the rotation **R** and the translation **t** from **E** is classical (Faugeras et al., 1987; Hartley, 1992; Longuet-Higgins, 1981; Tsai and Huang, 1984). We denote this algorithm by **FACTOR**.

An Iterative Solution. An alternative method is to use directly the error function that has been used to determine the fundamental matrix. It differs from the previous one by the fact that it uses again the measured points. In (Luong et al., 1993; Zhang et al., 1995) different parameterizations for this matrix have been proposed to take into account constraints on its structure and linear and non-linear criteria for its estimation were also considered. We have used the two error functions:

$$\sum_i \left(\mathbf{m}_i'^T \mathbf{F} \mathbf{m}_i\right)^2 \quad \text{subject to } \mathrm{Tr}(\mathbf{F}^T\mathbf{F}) = 1 \tag{21}$$

and

$$\sum_i \left\{ d\left(\mathbf{m}_i'^T, \mathbf{F}\mathbf{m}_i\right)^2 + d\left(\mathbf{m}_i^T, \mathbf{F}^T\mathbf{m}_i'\right)^2 \right\} \tag{22}$$

where d is a the Euclidean distance in the image plane between a point and a line.

We denote by **MIN-LIN**[7] the minimization of the error function (21) and by **MIN-DIST** the minimization of the error function (22). The knowledge of the intrinsic parameters allows us to minimize these error functions with respect to five motion parameters: we parameterize **T** by t_1/t_3, t_2/t_3 and **R** by the three-dimensional vector **r** whose direction is that of the axis of rotation and whose norm is equal to the rotation angle. Hence, we minimize with respect to **r** and **T** the error functions:

$$\sum_i \left(\mathbf{m}_i'^T \mathbf{A}^{-T} \mathbf{E} \mathbf{A} \mathbf{m}_i\right)^2$$

and

$$\sum_i \left\{ d\left(\mathbf{m}_i'^T, \mathbf{A}^{-T}\mathbf{E}\mathbf{A}\mathbf{m}_i\right)^2 + d\left(\mathbf{m}_i^T, (\mathbf{A}^{-T}\mathbf{E}\mathbf{A})^T\mathbf{m}_i'\right)^2 \right\}$$

where $\mathbf{E} = [\mathbf{T}]_\times \mathbf{R}$.

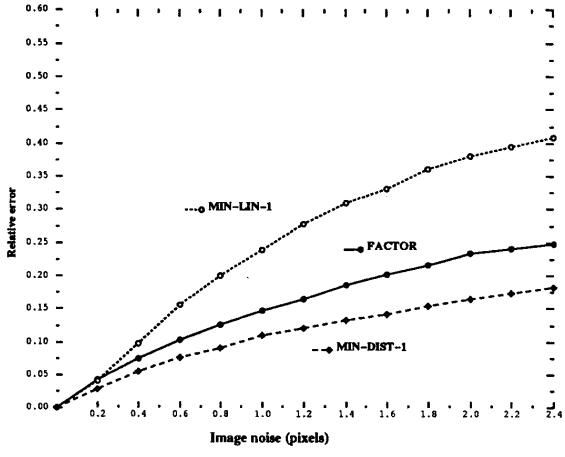

Figure 4. Relative error on the rotation, initialization with the exact displacement.

4.2. An Experimental Comparison

The Case of Exact Intrinsic Parameters. In the first comparative study, we suppose that the *exact* intrinsic parameters are known. The graphs have been obtained using 200 different displacements, and show the average relative error on the rotational and translational components, measured as $\frac{\|\Delta \mathbf{r}\|}{\|\mathbf{r}\|}$ and $\frac{\|\Delta \mathbf{t}\|}{\|\mathbf{t}\|}$, where **r** is a vector whose norm is the angle of the rotation, and whose direction gives the rotation axis. Since the non-linear methods require a starting point whose choice is important, we have considered the three possibilities:

1. the exact motion, to test the precision of the minimum (Figs. 4 and 5, Label 1).
2. the motion obtained by **FACTOR**, which is the realistic initialization (Figs. 6 and 7, Label 2).
3. an arbitrary motion: $\mathbf{r} = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2})^T$, $\mathbf{t} = (0, 0, 1)^T$, to test the convergence properties (Figs. 6 and 7, Label 3).

These indexes (1, 2, 3) are used as labels in the graphs plotted in Figs. 4–7. For example, **MIN-LIN-1** means that the corresponding curve shows the performance of the **MIN-LIN** error function when initialized with the exact motion.

The conclusions of the simulations are:

- The computation is more stable than the fundamental matrix computation. Motion computation is a less difficult problem.
- The rotational part is determined more precisely than the translational part.

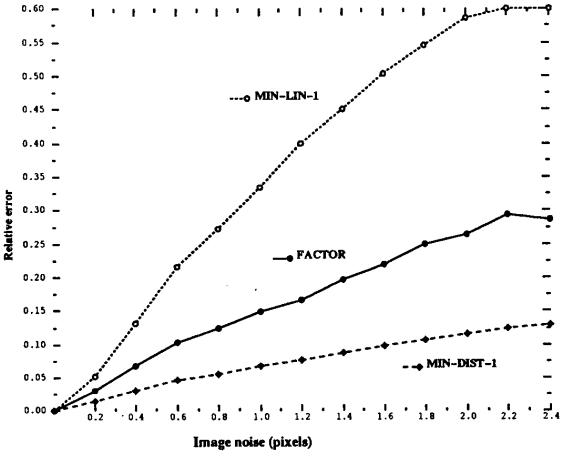

Figure 5. Relative error on the translation, initialization with the exact displacement.

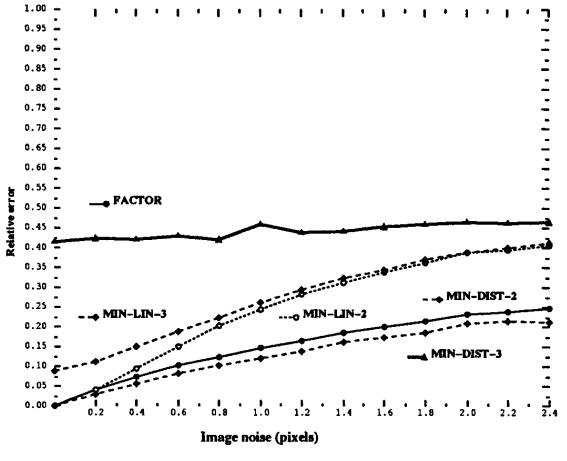

Figure 6. Relative error on the rotation, initialization with the results of **FACTOR** (2) and an arbitrary motion (3).

- The iterative method based on **MIN-DIST** is the most precise, but it is the most sensitive to the choice of the starting point.
- The results obtained by **MIN-DIST** and by **FACTOR** in the realistic case where **MIN-DIST** is initialized with the results of **FACTOR** are very close.

Note that even using **MIN-LIN**, the results are much more precise than those usually found by using a purely linear methods such as the eight-point algorithm (Fang and Huang, 1984; Tsai and Huang, 1984).

Sensitivity to Errors on the Intrinsic Parameters. Very few results are available concerning the sensitivity of motion and structure computations to errors on the intrinsic parameters (Kumar and Hanson, 1990). It is nevertheless an important issue, as it determines the precision of calibration that it is necessary to achieve to obtain a given precision on the three dimensional reconstruction, which is the final objective. We present here some experimental results which give an idea of the numerical values. All experiments were run with pixel noise levels varying from 0.2 to 1.8 pixels. Figure 8 represents the effects of the error on the location of the principal point. The exact principal point is at the center $(255, 255)$ of the image, and we have used for the computation of the motion principal points that were shifted from 20 to 200 pixels following a Gaussian law. Each point on the figure represents 100 trials.

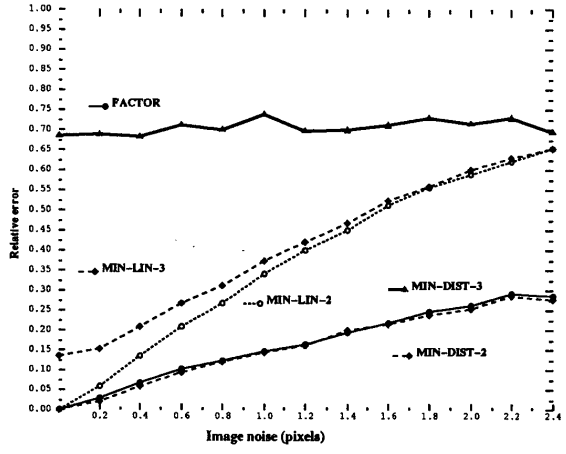

Figure 7. Relative error on the translation, initialization with the results of **FACTOR** (2) and an arbitrary motion (3).

Figure 9 represents the effects of the error on the scale factor, which has been similarly made vary from 2.5% to 25%. Among the numerous conclusions that can be drawn from the graphs, we would like to emphasize the following:

- The effects of the imprecision on intrinsic parameters are significant; however, until relatively large errors are reached (10% on the scale factors, several tens of pixels for the principal point), these effects are less significant than those due to noise (for example, if the image noise increases from 0.6 to 1.0 pixels).
- The sensitivity to errors on the principal point is less than the sensitivity to errors on the scale factor: in terms of relative errors, a 120 pixels shift of the principal point corresponds to a 50% relative error and has the same effects as a 25% relative error on the scale factors.
- The iterative criterion **MIN-DIST** is more sensitive to the error on the intrinsic parameters than the solution of **FACTOR**. This can be explained by the fact that the fundamental matrix, which is directly

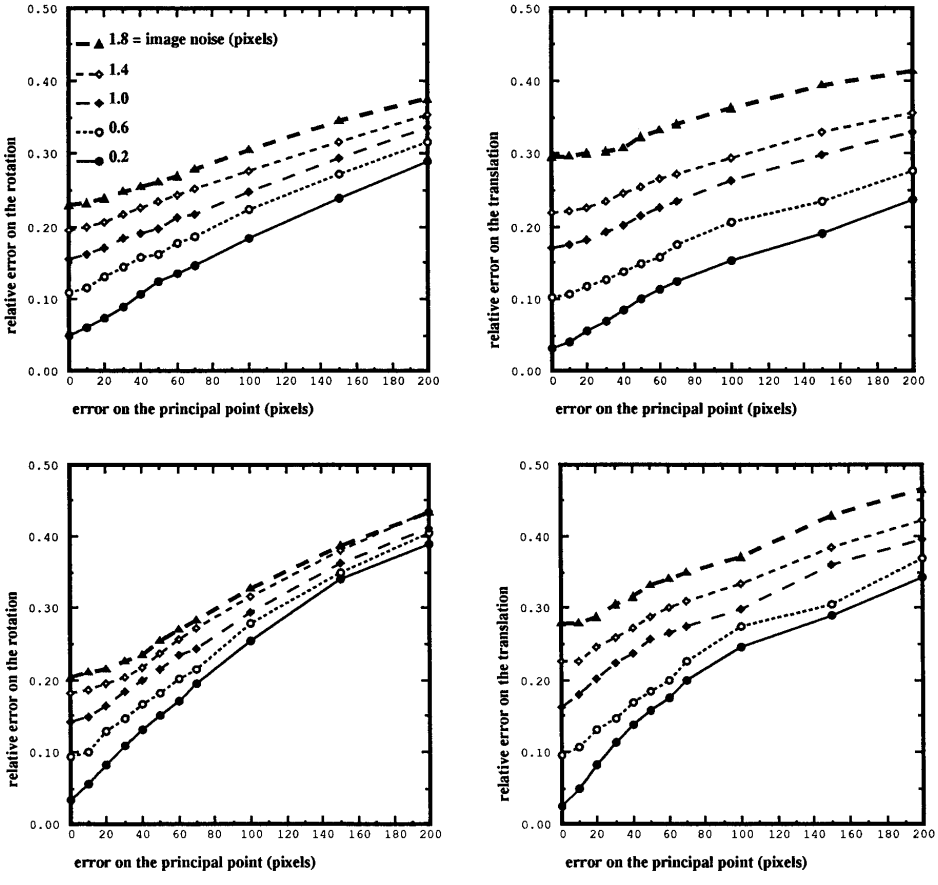

Figure 8. Sensitivity of motion computation to errors on the principal point. Top: **FACTOR**, bottom: **MIN-DIST**, left: rotation, right: translation.

used by **FACTOR** partially retains the information on the exact intrinsic parameters, whereas the iterative method compensates entirely the error on the intrinsic parameters by an error on the computed motion.

The conclusion of these two subsections is that we recommend in practice to use the error function **FACTOR** when the intrinsic parameters are not accurately known and **MIN-DIST** when they are, since it is less sensitive then **FACTOR** to pixel noise.

4.3. A Global Approach to Compute Simultaneously Calibration and Motion

Using a Single Displacement. A natural extension of the previous techniques is to minimize the error function (22) simultaneously with respect to the five motion parameters previously introduced and to the intrinsic parameters. Since we have seen that the most significant are α_u and α_v, we choose to estimate them keeping u_0 and v_0 constant and equal to some "reasonable" values. The relative errors obtained on the motion parameters are shown in Fig. 10. They are to be compared to Fig. 9, and to facilitate this comparison, we have also plotted on this figure the two curves obtained in Fig. 9 for the two extreme noise levels. This superposition makes it clear that the new method is much less sensitive to initial errors on the scale factors, but more sensitive to noise.

The final error on the motion are compensated by errors on the camera parameters, as seen in Fig. 11, which shows that the final error on the camera parameters depends mainly on the noise (the curves are almost horizontal lines), and not to much on the initial error on the parameters. The fact that the curves are not exactly horizontal is due to the fact that when the initial error on the camera parameters are above 25%, the algorithm has sometimes difficulties to retrieve the correct camera parameters, probably due to convergence problems. The graphs shows also that it is interesting

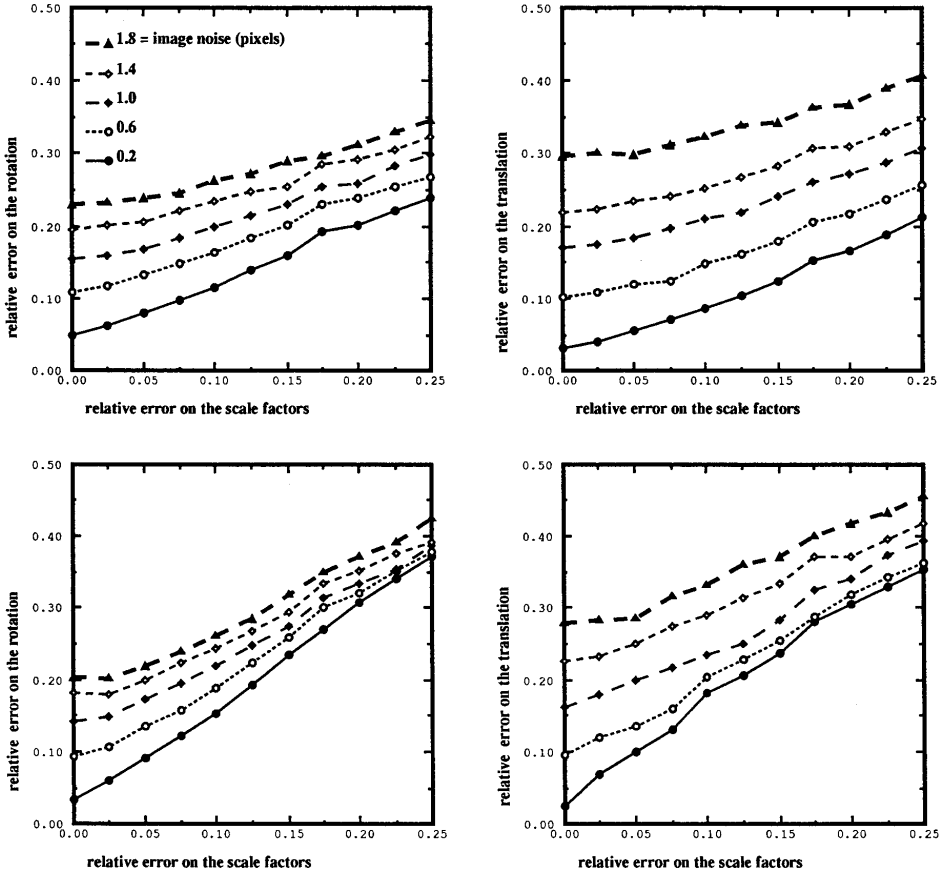

Figure 9. Sensitivity of motion computation to errors on the scale factors. Top: **FACTOR**, bottom: **MIN-DIST**, left: rotation, right: translation.

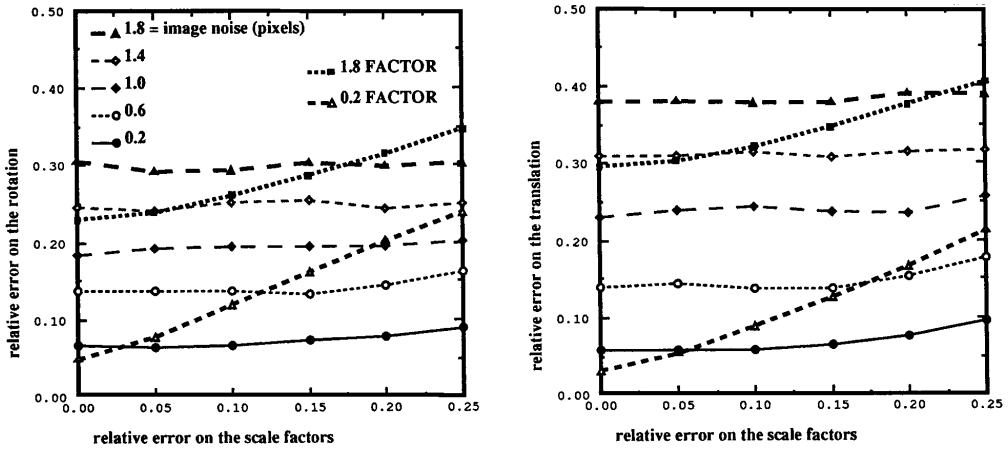

Figure 10. Sensitivity of motion computation to errors on the scale factors, in the case the scale factors are allowed to vary. Left: rotation, right: translation.

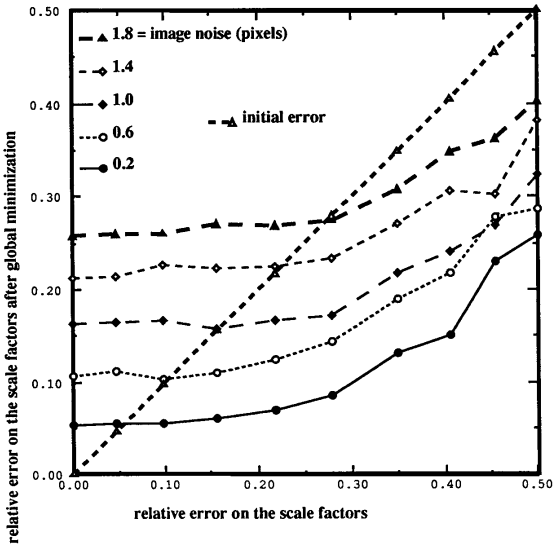

Figure 11. Final scale factors obtained when computing the motion.

to re-estimate the camera parameters only if the error on these parameters is sufficiently high, and this level depends again on the image noise.

Global Minimization Using Multiple Motions. If we have several camera displacements, then the previous approach can be used to estimate all the camera parameters and the displacements. Since the minimization is highly non-linear, and involves a large number of unknowns, to obtain convergence we need a good starting point, which can fortunately be obtained from the previous method. Let us summarize the new algorithm, which can accommodate N independent displacements ($N \geq 2$), and, for each displacement i, a minimum of eight correspondences $(\mathbf{m}_{ij}, \mathbf{m}'_{ij})_j$:

A Comparison. We now present some statistical simulation results to show that the new global method can significantly improve upon the results obtained by the methods based on the Kruppa equations (denoted by **KRUPPA**). The name **MOUV** designates the global method, initialized with the starting point already used in the previous section (800, 800, 255, 255). The name **MOUV-KRUPPA** designates the global method, initialized with the values obtained by **KRUPPA**. All names are followed by the number of displacements used, e.g., **KRUPPA2**. The image noise has the same meaning as previously, that is Gaussian noise added to pixel coordinates of point correspondences. In Fig. 12, each point represents 100 trials, obtained by varying the intrinsic parameters and the camera motions. We have represented the average error on the scale factors. We have given both the results with u_0 and v_0 fixed and varying.

Let us try to characterize the two methods **KRUPPA** and **MOUV**. The first stage for each method is identical: it is concerned with the determination of the fundamental matrices. Then in the second stage of determining the intrinsic parameters, **KRUPPA** use only these matrices, the rigidity constraint being used to *eliminate* the unknown motion parameters. Thus the method involves only the unknowns we try to compute, and allows for a semi-analytical solution, as well as for efficient iterative solutions. Contrary to this, in **MOUV**, it is the *form of the parameterization* which ensures that all the constraints are satisfied. Then we have to compute explicitly all the unknowns in the problem, and thus need a good starting point and more intensive computations. However, first the criterion takes into account more constraints, since it ensures the exact decomposability of each fundamental matrix $\mathbf{F}_i$ under the

Global computation of intrinsic parameters and motion

1. Compute the N fundamental matrices $\mathbf{F}_i$.
2. Compute an initial estimate of the intrinsic parameters (α_u, α_v, u_0, v_0), using one of the Kruppa methods described in Section (3).
3. Compute the N initial motions $(\mathbf{r}_i, \mathbf{t}_i)$ using **FACTOR**, from the $\mathbf{F}_i$ and the intrinsic parameters
4. Minimize, with respect to the $5N + 4$ variables (or $5N + 2$ if u_0, v_0 are taken as the image center) the error function:

$$\min_{\substack{\alpha_u,\alpha_v,u_0,v_0 \\ \mathbf{r}_i,\frac{t_{ix}}{t_{iz}},\frac{t_{iy}}{t_{iz}}\ i=1\cdots N}} \sum_{i=1}^{N} \sum_{j} d^2(\mathbf{m}'_{ij}, \mathbf{A}^{-1T}[\mathbf{t}_i]_\times \mathbf{R}_i \mathbf{A}^{-1}\mathbf{m}_{ij}) + d^2\big(\mathbf{m}_{ij}, \mathbf{A}^{-1T}\mathbf{R}_i^T[\mathbf{t}_i]_\times \mathbf{A}^{-1}\mathbf{m}'_{ij}\big) \qquad (23)$$

   where d is the Euclidean point-line distance, $\mathbf{R}_i = e^{[\mathbf{r}_i]_\times}$, and $\mathbf{A} = \begin{bmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}$

5. Perform again stage 3 with the new intrinsic parameters (optional).

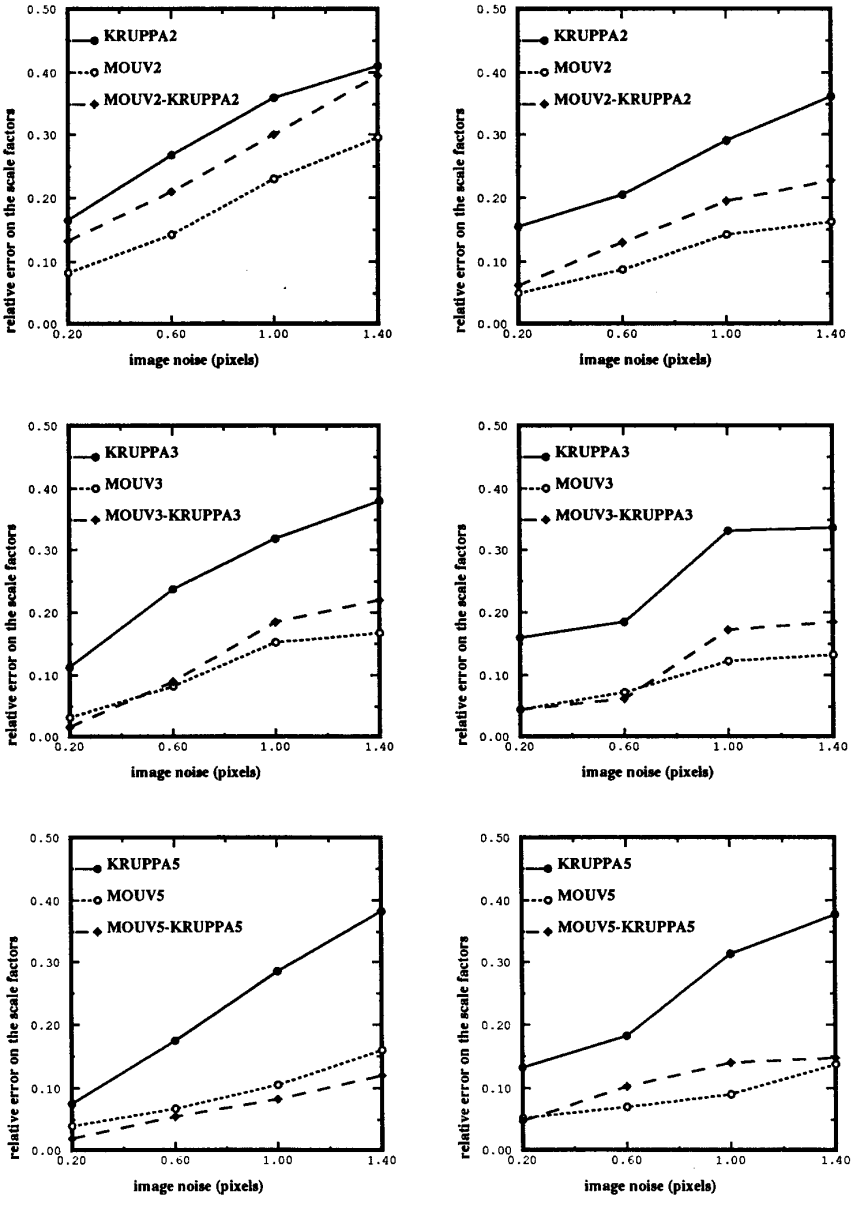

Figure 12. Comparison of the Kruppa based self-calibration method with the motion-based global self-calibration method. Left: 4 parameters estimated (variable principal point). Right: 2 parameters estimated (fixed principal point).

form $\mathbf{A}^{-1T}\mathbf{E}_i\mathbf{A}^{-1}$, with an unique intrinsic parameters matrix. Thus it achieves a minimal parameterization of the unknowns. In the **KRUPPA** approach, the fundamental matrices obtained verify further constraints, which are precisely the existence of solutions for the Kruppa equations, and these constraints cannot be enforced at the first stage of the computation. Second, the criterion uses directly more information. This explains why we obtain more precise results.

4.4. An Evaluation of the Methods Based on the Motion

In this section, we have presented an alternative method for self-calibration, which computes simultaneously the intrinsic parameters and the camera motion.

As a preliminary, we have first studied the computation of the motion parameters in the context of self-calibration. One finding is that although in a classical

context where the camera parameters are known accurately, the non-linear minimization techniques provide the most accurate results for the motion parameters, the best method in our context is the decomposition of essential matrix method. This method is very fast and not very sensitive to the errors on the camera parameters.

Once an estimate of the motion has been obtained this way, we can simultaneously refine the camera and motion parameters. So far this method has proved to be the most reliable, and gives better results than the methods based on the Kruppa equations. Although its principle is very simple, it nevertheless depends on the availability of a starting point, and the methods presented in the previous section are perfectly adequate for this purpose, since some of them do not even need an initialization.

5. Experimental Results with Real Data

5.1. Self-Calibration of a Camera

We use three images taken by a camera from different positions. The camera is a Pulnix CCD camera. The CCD has a size of 6.4 mm $\times$ 4.8 mm, and the lens has a focal of 8 mm, resulting in an horizontal field of view of 43°. In order to make comparisons possible with the standard calibration method, we have performed displacements in such a way that the calibration grid remains always visible. We use between 20 and 30 corners, which are extracted with a sub-pixel accuracy, semi-automatically, by the program of Deriche and Blaszka (1993). Correspondence is performed manually. It should be noted that the corresponding points between pairs of images are different, that is, points need not be seen in the three views. Figure 13 shows the detected points of interest matched between image 1 and image 2.

Note that only few of the points are on the calibration grid. The standard calibration is performed on each image, using the algorithm of Robert (1993), which is a much improved version of the linear method of Faugeras and Toscani (1986). From the projection matrices obtained by this algorithm, the three fundamental matrices $\mathbf{F}_{12}$, $\mathbf{F}_{23}$, $\mathbf{F}_{13}$ are computed and used as a reference for the comparisons with our algorithm which computes the fundamental matrices from the point matches. The resulting epipoles are shown in Table 6.

Table 6. Results of the fundamental matrix estimation.

	From the grid				Estimated				RMS	
	e_x	e_y	e'_x	e'_y	e_x	e_y	e'_x	e'_y	Points	Grid
1-2	−222.4	181.0	−466.9	167.5	−200.0	185.8	−447.5	170.1	0.36	0.76
2-3	2226.9	−1065.1	−2817.9	1646.6	2708.5	−1380.1	−2099.6	1315.5	0.31	0.31
1-3	654.4	−288.8	1114.7	−715.6	680.2	−321.7	1230.9	−842.2	0.26	0.54

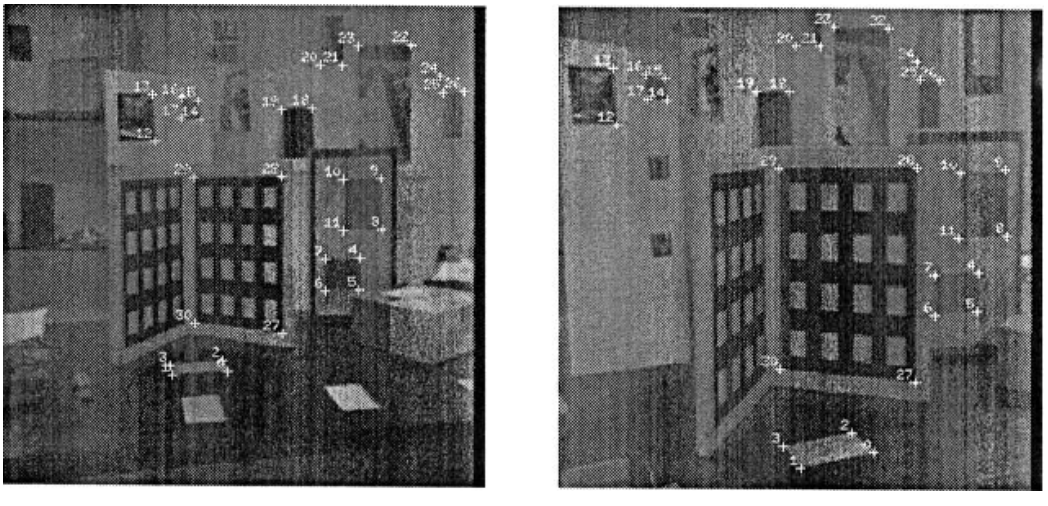

Figure 13. A pair of images with the detected corners superimposed.

Table 7. Results of the intrinsic parameters estimation.

Method	α_u	α_v	u_0	v_0	$\theta - \frac{\pi}{2}$
Grid, image 1	657	1003	244	256	−2.05e-06
Grid, image 2	664	1015	232	257	−7.47e-07
Grid, image 3	639	980	252	249	−2.60e-06
Average, [s.d.]	653 [10]	999 [14]	242 [8]	254 [3]	
Kruppa polynomial	639	982	258	341	−6.11e-03
Kruppa iterative	640	936	206	284	−0.07
Kruppa iterative (center)	681	985	255	255	
Average, [s.d.]	653 [19]	967 [22]	239 [23]	293 [35]	

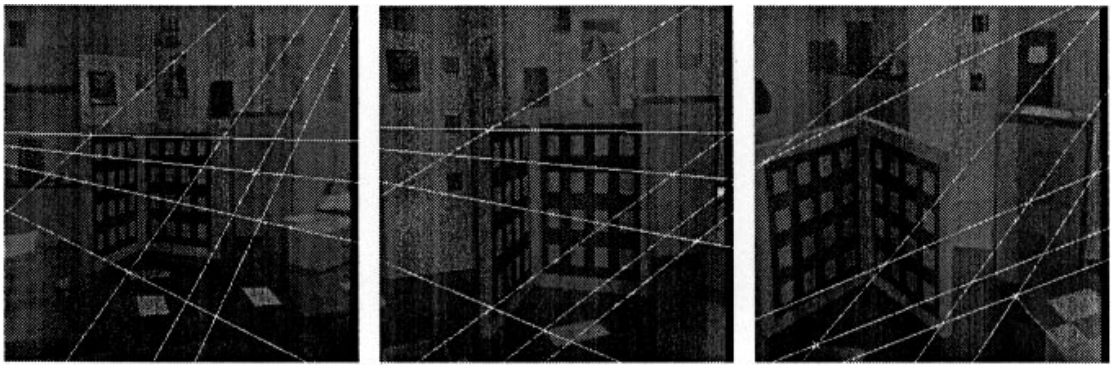

Figure 14. A triplet of images with some estimated epipolar lines superimposed.

It can be seen that the estimation is quite precise. We have given two values of the RMS error, which represents the average distance of corresponding points to epipolar lines. The first one (labeled points) is computed over the detected points which were used to estimate the fundamental matrices. The low value (one third of a pixel) confirms the validity of our linear distortion-free model (see remark in Section 2.1), as well as the accuracy of the corner detection process. The second value of the RMS (grid) is computed over the 128 corners of the little white squares on the calibration grid, which were used for model-based calibration. Since these points were not used at all to estimate the fundamental matrices, this provides appropriate control values. As expected, the RMS with the control points is sometimes higher than the RMS with the data points, but the values remains bellow one pixel. Some epipolar lines obtained with points that are seen in the three images are shown in Fig. 14 to illustrate the quality of the estimated epipolar geometry. The cameras intrinsic parameters are then computed from the fundamental matrices. We show in Table 7 the intrinsic parameters obtained by the standard calibration method using each of the three images, and the results of our method, with the polynomial method (Section 3.2) and the iterative method (Section 3.3) used to compute all the parameters, or just the scale factors, starting from the previous value.

The scale factors are determined with a good accuracy, however, this is not the case for the coordinates of the principal point. Thus the best is to assume that it is at the center of the image. Note that the intrinsic parameters computed from the standard calibration method show a fair amount of variability among views. We have then compared in the Table 8 the camera motion obtained directly from the projection matrices given by the classic calibration procedure, and the estimation by performing the decomposition of the fundamental matrices already obtained, and using the camera parameters obtained by the self-calibration method. The table shows the relative error on the rotation angle α, the angular error θ_r on the rotation axis and θ_t on the direction of translation. It can be seen that the estimation is accurate.

Table 8. Results of the camera motion estimation. Rotations are represented by the rotation vector, translations by their unit direction.

Motion	r_x	r_y	r_z	t_x	t_y	t_z	$\frac{\Delta\alpha}{\alpha}$	θ_r	θ_t
1-2 grid	0.01175	−0.2117	−0.01785	−0.7290	−0.06831	0.6809			
Estimated	0.01843	−0.2110	−0.01961	−0.7239	−0.06102	0.6871	0.0005	1.8	0.62
2-3 grid	0.1900	0.4526	0.1211	−0.9395	0.2779	0.1999			
Estimated	0.1915	0.4682	0.1279	−0.9209	0.2896	0.2608	0.032	0.61	3.7
1-3 grid	0.2007	0.2533	0.07876	0.6976	−0.5041	0.5090			
Estimated	0.01306	−0.2145	−0.01405	−0.7371	−0.05872	0.6731	0.10	0.98	3.0

Table 9. Parameters obtained with a zoom camera.

Focal	Method	α_u	α_v	u_0	v_0	$\theta - \frac{\pi}{2}$	$\frac{\alpha_u}{\alpha_v}$
9	GRID	481.31	711.54	248.57	260.97	10^{-7}	.6764
	SELFCALIB	503.49	760.71	250.24	282.67		.6618
12	GRID	642.45	950.37	248.30	263.31	-5.10^{-7}	.6759
	SELFCALIB	636.12	921.36	201.52	338.89		.6904
20	GRID	1036.38	1539.6	252.43	272.53	7.10^{-8}	.6731
	SELFCALIB	1208.83	1838.48	251.93	200.58		.6575
30	GRID	1573.20	2330.953	207.98	210.35	4.10^{-7}	.6749
	SELFCALIB	2047.61	3063.94	249.678	198.463		.6682

5.2. *Varying the Focal Length*

We have applied the method to a camera with a variable focal length. The camera was fitted with a Canon zoom lens J8x6B4 which covers focals from 6 mm to 48 mm. The size of the CCD is 4.8 mm × 3.6 mm. In order to avoid camera distortion, we have not used the focals shorter than 9 mm. The results are shown in Table 9. It allows us to notice that the best results are obtained for short focal lengths, which yield large fields of views. Although the focal length is overestimated by the method for large values, we can notice that the computed aspect ratio is quite consistent over the whole focal range.

5.3. *Reconstructions from a Triplet of Uncalibrated Images Taken by a Camera*

We now show an example of reconstruction using structure from motion with three uncalibrated views. The approach is to use the global minimization approach presented in Section 4.3, with the variant to account for trinocular constraints.

We have tested the precision of reconstruction of our algorithm using triplets of images of a standard photogrammetric calibration pattern which were communicated to us for testing by the commercial photogrammetry company *CHROMA*, of Marseille, France. Coordinates of 3D reference points are available, which allows us to assess quantitatively the error in reconstruction from the uncalibrated images. The triplet used in this experiment is shown in Fig. 15.

The points of interest are the light dots and have been located and matched manually[8]. Note that the scale factors found $\alpha_u = 1859.47$, $\alpha_v = 2520.79$ correspond to a rather long focal length, which is not very favorable, and that among the three motions between pairs of images, the motion 2-3, whose translation vector was found to be $\mathbf{t}_{23} = (-1.186, 0.6623, -0.0857)^T$, is nearly parallel to the image plane, a defavorable configuration, as shown in (Luong, 1992; Luong and Faugeras, 1995). However, the epipolar geometry found from the three projection matrices obtained by self-calibration is fairly coherent, as illustrated in Fig. 16, which shows a zoom with the epipolar lines of one the point of interest.

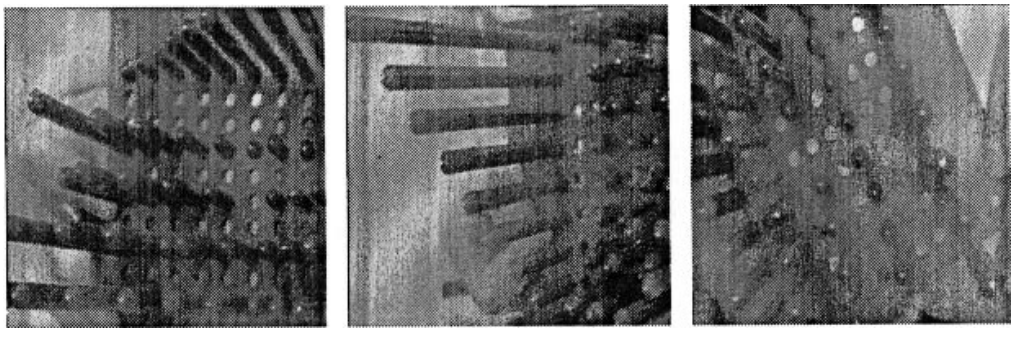

Figure 15. The triplet of images of the photogrammetric object.

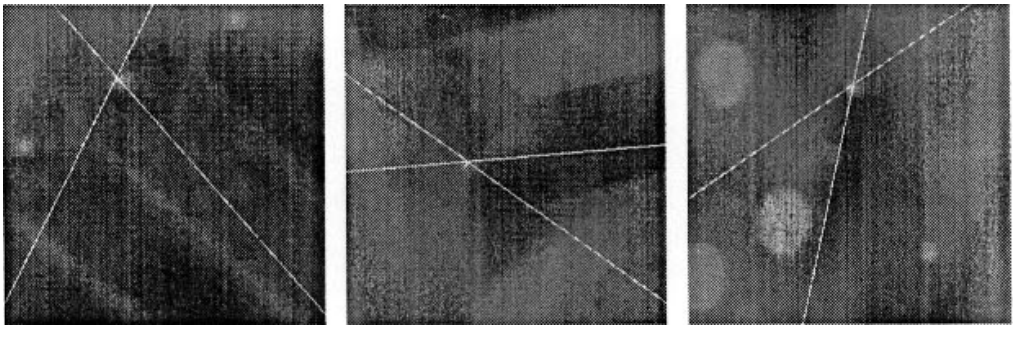

Figure 16. Zoom on the photogrammetric triplet, showing corresponding epipolar lines.

We have then performed a 3D trinocular reconstruction from the matched points, using our computed projection matrices as input for a the classical reconstruction algorithm of Deriche et al. (1992). The 3D points are obtained in the coordinate system associated with one of the cameras, since we can reconstruct only up a similarity with the self-calibration technique. Thus in order to compare the reconstruction with the reference data, we have computed the best similarity which relates the two sets of 3D points, using an algorithm of Zhang. After applying this similarity to the initial reconstruction, the final average error in 3D space with this sequence is 2 millimeters[9]. A sample of coordinates of reconstructed points are shown in Table 10, units being in millimeters.

Table 10. Comparison of the 3D reconstruction from self-calibration with reference points.

Reference points			Reconstructed points		
X	*Y*	*Z*	*X*	*Y*	*Z*
−56.3	0.38	90.1	−55.5	−2.28	89.1
−69.7	0.33	110.1	−69.6	−3.02	108.3
−41.8	30.0	40.1	−40.9	29.7	40.6
−28.2	49.8	90.0	−26.5	49.3	89.2
−70.0	30.0	3.5	−69.8	30.4	3.4
−112.0	70.2	90.1	−113.8	70.5	88.1
−69.5	89.7	90.0	−69.6	90.8	88.9

6. Conclusion

We have presented a general framework to perform the self-calibration of a camera. The basic idea is that the only information which is needed to perform calibration are point correspondences between images taken from various viewpoints. This is contrast with all standard calibration methods. As a side effect of the calibration procedure, we can also estimate the relative displacements between the cameras and the structure of the scene. The algorithms which arise from this study are the most general possible, in the sense that they do not require:

- any model of the observed objects, or any 3D coordinates,
- any knowledge of the camera motion, which can be entirely general, with the exception of a few

degenerate cases, and can be computed as a byproduct of the method,
- any initial guess about the values of the camera parameters, or any restrictive model of these parameters, which describe the most general projective camera model.

Thus, of the four pieces of information used in 3D vision (calibration, motion, structure, correspondences), our method needs only one input and produces three outputs, whereas the other algorithms need at least two inputs or produce at most two outputs, as shown in the table below:

Paradigm	Camera parameters	Correspondences	Rigid displacement	3D Structure
Structure from motion	Input	Input	Output	Output
Stereovision	Input	Output	Input	Output
Model-based calibration	Output	Input	Not used	Input
Calibration from motion	Output	Input	Input	Not used

The problem of on-line calibration is now becoming very important in the framework of active vision, where optical parameters such as focus, aperture, zoom, and vergence are constantly changing, making the use of classic calibration techniques impossible. Another critical framework is digital image libraries which are usually built from images for which no calibration data is available. Thus a number of researchers have recently investigated self-calibration techniques. However, with one exception, all of them have put more limitations on their methods than we did, by adding supplementary constraints, such as an initial knowledge of camera parameters which are then only updated (Crowley et al., 1993), or restriction on the camera motions (Basu, 1993; Dron, 1993; Du and Brady, 1993; Hartley, 1994b; Viéville, 1994). When the camera motion is exactly known in some reference frame, then these methods should be rather called "calibration from motion" than self-calibration, where motion *and* calibration are estimated. However, one of the most reasonable restriction seems to be a partial control of the motion, which may be performed by a robotic head. In this context, one of the most general work is that of Viéville (1994) where the only additional assumption is the fact that the motion is a fixed-axis rotation, something well-suited to robotics heads. Another approach, which is complementary to the one described in this paper, is to use a stationary camera (Hartley, 1994b; Luong and Viéville, 1993). More precise and robust results can be obtained, since the problem is more constrained.

The only equivalent approach has been recently[10] presented by Hartley (1994a). There are a number of similarities in the steps of the algorithm, although each step is quite differently done. The middle steps are in both methods non-iterative computations mixed with small-scale optimizations based on motion parameters in order to find an initialization for the camera parameters. The first step, and the last step consists in exploiting directly the point correspondences. The main difference here is that Hartley's method it is based on a bundle adjustment technique, whereas ours is based on the epipolar geometry. Therefore, Hartley's algorithm enforces more geometric constraints, but requires correspondences across multiple views, and a large scale non-linear minimization which is not successful with the minimum number of views. Experimental comparisons remain to be done in order to quantify the eventual gain in precision.

Although we have shown using experiments with real images that our self-calibration method can be accurate enough to provide useful 3D metric descriptions, and that the results are often of a similar quality to the ones obtained by a traditional method, it must be admitted that the method has presently its own constraints: not all types of displacements yield stable results, and, as in all calibration procedures, precise image points localization and reliable correspondences are necessary. The precision of our method might compare very defavorably with those obtained using photogrammetric techniques, but one has to remember that our method can work with only the theoretical minimal amount of data, and although the reconstructions are not very precise, they are usable for most robotics tasks, and are certainly better than the total absence of metric information. Only three images are required, and there is no need to have correspondences across three views. By using more images, which has not been done in the examples of this paper, results would improve.

Natural extensions of this work are to investigate the geometry of a system of three cameras, since our formulation does not take into account trinocular constraints at the projective level, but only at the Euclidean

level. Using a third view should also enable to use lines, which are usually more stable primitive than points. It can be expected that the resulting algorithm will have nicer robustness properties. Another idea, which is important in the framework of active vision, is to study the case of parameters which are allowed to change over time. The framework that has been laid out in this paper could prove to be a useful starting point for these studies which would hopefully result in more truly autonomous vision systems.

Appendix A: Equivalence of the Trivedi Equations and the Huang-Faugeras Constraints

We now show that the three Trivedi equations are equivalent to the Huang and Faugeras conditions. Let first suppose that we have (13). It follows immediately that $\det(\mathbf{E}\mathbf{E}^T) = 0$, and thus the first condition $\det(\mathbf{E}) = 0$ is satisfied. Adding T_{12}, T_{13} and T_{23}, yields:

$$4\left(S_{12}^2 + S_{13}^2 + S_{23}^2\right) + S_{11}^2 + S_{22}^2 + S_{33}^2 - 2(S_{11}S_{22} + S_{22}S_{33} + S_{33}S_{11}) = 0$$

Since the matrix $\mathbf{S}$ is symmetrical, the first term can be replaced by: $4(S_{12}S_{21} + S_{13}S_{31} + S_{23}S_{32})$, and a simple calculus shows that it is identical to the second Huang-Faugeras condition:

$$\text{trace}^2(\mathbf{S}) - 2\text{trace}(\mathbf{S}^2) = 0$$

Let then suppose that the Huang-Faugeras conditions are satisfied. They are equivalent to the fact that the matrix $\mathbf{E}$ has a zero singular value and two non-zero equal singular values σ. By definition, there exists an orthogonal matrix Θ such as:

$$\mathbf{S} = \mathbf{E}\mathbf{E}^T = \Theta \begin{bmatrix} 0 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix} \Theta^T$$

This matrix equality can be expanded as:

$$\mathbf{S}^2 = (\Theta_{i2}\Theta_{j2} + \Theta_{i3}\Theta_{j3})_{1 \le i,j \le 3}$$

Since $\mathbf{\Theta}$ is orthogonal:

$$\Theta_{i2}\Theta_{j2} + \Theta_{i3}\Theta_{j3} = \begin{cases} -\Theta_{i1}\Theta_{j1} & \text{if } i \neq j \\ 1 - \Theta_{i1}^2 & \text{if } i = j \end{cases}$$

The diagonal element $1 - \Theta_{11}^2$ (resp. $1 - \Theta_{21}^2, 1 - \Theta_{31}^2$) can be rewritten $\Theta_{31}^2 + \Theta_{21}^2$ (resp. $\Theta_{11}^2 + \Theta_{31}^2$, $\Theta_{21}^2 + \Theta_{11}^2$), which shows that $\mathbf{S}$ has exactly the form (13).

Appendix B: Equivalence of the Huang-Faugeras Constraints and the Kruppa Equations

Let us make a change of retinal coordinate system in each of the two retinal planes, so that the new fundamental matrix is diagonalised. One way to see that it can always be done is to use the singular value decomposition : there exists two orthogonal matrices $\mathbf{\Theta}$ and $\mathbf{\Delta}$ such that $\mathbf{F} = \mathbf{\Delta}\mathbf{\Lambda}\mathbf{\Theta}^T$. If we use matrix $\mathbf{\Theta}$ to change retinal coordinates in the first retina and matrix $\mathbf{\Delta}$ to change retinal coordinates in the second retina, the new intrinsic parameters matrices are $\mathbf{A} = \mathbf{A}_0\mathbf{\Theta}$ and $\mathbf{A}' = \mathbf{A}_0\mathbf{\Delta}$ in the first and second retina, respectively. If the epipolar constraint in normalized coordinates $\mathbf{m}$ and $\mathbf{m}'$ was:

$$\mathbf{m}'^T \mathbf{A}_0^{-1T} \mathbf{F} \mathbf{A}_0^{-1} \mathbf{m} = 0$$

with the new coordinate systems, we have:

$$\mathbf{p}'^T \mathbf{A}'^{-1T} \mathbf{\Lambda} \mathbf{A}^{-1} \mathbf{p} = 0$$

Thus it is possible, *provided we allow the two cameras to be different*, to consider that $\mathbf{F}$ is in diagonal form:

$$\mathbf{F} = \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \mu & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \text{(B1)}$$

where $\lambda \neq 0$ and $\mu \neq 0$ since we know that a fundamental matrix must be of rank two. Let's use the Kruppa notation:

$$\mathbf{K} = \begin{pmatrix} -\delta_{23} & \delta_3 & \delta_2 \\ \delta_3 & -\delta_{13} & \delta_1 \\ \delta_2 & \delta_1 & -\delta_{12} \end{pmatrix}$$

Using (B1) we obtain easily the epipoles $\mathbf{e} = \mathbf{e}' = (0, 0, 1)^T$ and then, after some algebra, the Kruppa equations:

$$\lambda\delta_3\delta'_{23} + \mu\delta_{13}\delta'_3 = 0 \qquad (E_1)$$
$$\lambda\delta_{23}\delta'_3 + \mu\delta_3\delta'_{13} = 0 \qquad (E_2)$$
$$\lambda^2\delta_{23}\delta'_{23} - \mu^2\delta_{13}\delta'_{13} = 0 \qquad (E_3)$$

with, $\mathbf{l}_1^T$, $\mathbf{l}_2^T$, $\mathbf{l}_3^T$ being the row vectors of $\mathbf{A}$ (similar primed notations are used for the second retina):

$$\begin{aligned}\delta_3 &= \langle \mathbf{l}_1, \mathbf{l}_2 \rangle \\ \delta_{13} &= -\|\mathbf{l}_2\|^2 \\ \delta_{23} &= -\|\mathbf{l}_1\|^2\end{aligned} \tag{B2}$$

Note that although we use for convenience the three Kruppa equations, only two of them are independent, since we have for instance the relation:

$$\lambda\delta_{23}E_1 - \mu\delta_{13}E_2 = \delta_3 E_3 \tag{B3}$$

Let now express the condition $f(\mathbf{E}) = 0$. Since: $\mathbf{E} = \mathbf{A}'^T\mathbf{F}\mathbf{A}$, some algebra (done partially using the symbolic computation program MAPLE), leads to:

$$\begin{aligned}f(\mathbf{E}) = -\frac{1}{2}&((\lambda^2\delta_{23}\delta'_{23} - \mu^2\delta_{13}\delta'_{13})^2 \\ &+ 2\lambda\mu(\lambda\delta_3\delta'_{23} + \mu\delta_{13}\delta'_3)(\lambda\delta_{23}\delta'_3 + \mu\delta_3\delta'_{13}))\end{aligned}$$

or

$$f(\mathbf{E}) = -\frac{1}{2}\left(E_3^2 + 2\lambda\mu E_1 E_2\right)$$

It is then clear that if the Kruppa equations are satisfied, then $f(\mathbf{E}) = 0$. Let now prove the inverse implication.

In the case where $\delta_3 \neq 0$, the previous equation can be rewritten, using (B3):

$$(\lambda\delta_{23}E_1 - \mu\delta_{13}E_2)^2 + 2\lambda\mu E_1 E_2 \delta_3^2 = 0 \tag{B4}$$

Thus:

$$\lambda^2\delta_{23}^2E_1^2 + \mu^2\delta_{13}^2E_2^2 = 2\lambda\mu E_1E_2\left(\delta_{13}\delta_{23} - \delta_3^2\right) \tag{B5}$$

According to the definitions (B2) of δ_3, δ_{13}, δ_{23}, the Schwartz inequality implies that $\delta_{13}\delta_{23} - \delta_3^2$ is superior or equal to zero. If it is zero, one can obtain from (B5) that $\delta_{23}E_1 = \delta_{13}E_2 = 0$. Since $\delta_{13}\delta_{23} = \delta_3^2 \neq 0$, it follows $E_1 = E_2 = 0$. If it is strictely positive, then $2\lambda\mu E_1E_2 \geq 0$. The Eq. (B4) is the sum of two positive terms, thus they have to be simultaneously zero, thus $E_1E_2 = 0$ and $E_3 = 0$.

The only special case which remains is $\delta_3 = 0$. The Kruppa equations are then in the simple form:

$$\mu\delta_{13}\delta'_3 = \lambda\delta_{23}\delta'_3 = \lambda^2\delta_{23}\delta'_{23} - \mu^2\delta_{13}\delta'_{13} = 0$$

which is equivalent to:

$$\begin{cases}\delta'_3 = 0 \\ \lambda^2\delta_{23}\delta'_{23} - \mu^2\delta_{13}\delta'_{13} = 0\end{cases}$$

or

$$\begin{cases}\delta'_3 \neq 0 \\ \delta_{13} = \delta_{23} = 0\end{cases}$$

and to:

$$f(\mathbf{E}) = 2\lambda^2\mu^2\delta_{13}\delta_{23}\delta'^2_3 + (\lambda^2\delta_{23}\delta'_{23} - \mu^2\delta_{13}\delta'_{13})^2 = 0$$

Appendix C: Independence of the Kruppa Equations Obtained from Three Images

The first two displacements are:

$$\mathbf{R}_1 = \begin{bmatrix}1 & 0 & 0\\ 0 & 0 & -1\\ 0 & 1 & 0\end{bmatrix} \quad \mathbf{t}_1 = \begin{bmatrix}1\\2\\1\end{bmatrix}$$

$$\mathbf{R}_2 = \begin{bmatrix}0 & 1 & 0\\ -1 & 0 & 0\\ 0 & 0 & 1\end{bmatrix} \quad \mathbf{t}_2 = \begin{bmatrix}2\\0\\-1\end{bmatrix}$$

The displacement obtained by composition of $\mathbf{D}_1$ and $\mathbf{D}_2$, in the coordinate system of the first camera is:

$$\mathbf{R}_3 = \mathbf{R}_1\mathbf{R}_2 = \begin{bmatrix}0 & 1 & 0\\ 0 & 0 & -1\\ -1 & 0 & 0\end{bmatrix}$$

$$\mathbf{t}_3 = \mathbf{R}_1\mathbf{t}_2 + \mathbf{t}_1 = \begin{bmatrix}3\\3\\1\end{bmatrix}$$

If we take as intrinsic parameter matrix $\mathbf{A}$ the identity matrix, the fundamental matrices are identical to the essential matrices. By chosing the normalization $\delta_{12} = 1$, the six Kruppa equations obtained are shown in Table 11. A solution of the system of equations E_1, E'_1, E_2, E'_2 obtained from the displacements $\mathbf{D}_1$ and $\mathbf{D}_2$ is:

$$\delta_1 = 0 \;\; \delta_2 = -\frac{1}{2} \;\; \delta_3 = 1 \;\; \delta_{13} = -4 \;\; \delta_{23} = 0$$

Substituting these values into the equations obtained from $\mathbf{D}_3$ yields: $E_3 = -27$, $E'_3 = 19$, thus we have verified that these equations are independant from the previous ones.

Table 11. The six Kruppa equations.

$$
\begin{aligned}
E_1 = {} & 3\,\delta_1 - 2 + 6\,\delta_3\,\delta_{13} + 9\,\delta_3 + 4\,\delta_1\,\delta_3 - 7\,\delta_2\,\delta_{13} \\
& + 2\,\delta_2 + 12\,\delta_1\,\delta_2 + 3\,\delta_1\,\delta_{13} + 2\,{\delta_{13}}^2 \\
E_1' = {} & 3\,\delta_{23} - 3\,\delta_{13}\,\delta_{23} + 8\,\delta_1\,\delta_{23} + 1 + \delta_1 - \delta_2\,\delta_{13} \\
& - 4\,\delta_2 - 4\,\delta_1\,\delta_2 - 4\,\delta_3\,\delta_{13} - \delta_3 + 4\,\delta_1\,\delta_3 \\
& - {\delta_{13}}^2 + \delta_1\,\delta_{13} \\
E_2 = {} & 2\,\delta_3\,\delta_{23} + 16\,\delta_3 - 8\,\delta_2\,\delta_3 + 4\,\delta_2\,\delta_{23} + 16\,\delta_2 \\
& - 16\,{\delta_2}^2 + 4\,\delta_1\,\delta_{13} + 16\,\delta_1 - 16\,{\delta_1}^2 \\
& + 2\,\delta_3\,\delta_{13} - 8\,\delta_1\,\delta_3 \\
E_2' = {} & {\delta_{23}}^2 + 4\,\delta_{23} - 4\,\delta_2\,\delta_{23} - {\delta_{13}}^2 - 4\,\delta_{13} + 4\,\delta_1\,\delta_{13} \\
E_3 = {} & 6\,\delta_{23} + 6\,\delta_3 + 18\,\delta_{23}\,\delta_3 + 12\,\delta_3\,\delta_{13} + 36\,{\delta_3}^2 \\
& + 18\,\delta_{23}\,\delta_2 + 36\,\delta_2\,\delta_{13} + 36\,\delta_2\,\delta_3 - 6\,\delta_{23}\,\delta_1 \\
& - 12\,\delta_1\,\delta_{13} + 18\,\delta_2 - 36\,\delta_1\,\delta_2 \\
& - 6\,{\delta_{13}}^2 - 18\,\delta_1 + 36\,{\delta_1}^2 \\
E_3' = {} & 9\,{\delta_{23}}^2 + 9\,\delta_{23}\,\delta_{13} + 18\,\delta_{23}\,\delta_3 + \delta_{23} - 9\,\delta_{13} \\
& + 2\,\delta_3 + 6\,\delta_{23}\,\delta_2 + 6\,\delta_2\,\delta_{13} + 12\,\delta_2\,\delta_3 - {\delta_{13}}^2 \\
& - 4\,\delta_1\,\delta_{13} - 9 + 12\,\delta_1 + 12\,{\delta_1}^2
\end{aligned}
$$

Acknowledgments

The authors would like to thank R. Deriche, S. Maybank, T. Papadopoulo, T. Viéville, and Z. Zhang for useful discussions and partial contributions to this work, T. Blaszka and B. Bascle for providing us with point of interest detectors, L. Robert for helping us with his calibration and stereo software, and H. Mathieu for making image acquisition possible.

Notes

1. Since a camera is characterized by its intrinsic parameters, this means that we assume that intrinsic parameters remain constant during the displacements. In the opposite case, the problem we would have to deal with would be the same as with multiple different cameras.
2. This relative distance had to be chosen, because the orders of magnitude of each component are very different.
3. In our implementation, we chose to compute the mean value, and to discard iteratively the solutions whose distance to the mean values are above a certain threshold. The final solution is obtained as the mean value of the retained solutions, and an estimate of covariance is obtained by computing their standard deviation.
4. The results improve if one considers more displacements. See next paragraph.
5. As it is a classical tool in computer vision, we do not give details on the filter itself, and rather invite the interested reader to read the classical references (Jazwinsky, 1970; Maybeck, 1979), or the more practical presentations which can be found in (Ayache, 1990; Faugeras, 1993 and Zhang and Faugeras, 1992).
6. The seemingly inferior results come from the fact that there was no requirements on these experiments on the minimum number of point matches generated, and thus often very few points have been used, in contrast with the previous experiments, where we started with at least 30 points.
7. Although it is not a linear method, but a non-linear method based on the same error measure than the linear criterion for the computation of the fundamental matrix.
8. A snake-based ellipse localization program due to B. Bascle, has also been tried.
9. This is typical, more precise results have been sometimes achieved.
10. The methods described in this paper were first described in (Luong, 1992). Parts of the results were presented in (Faugeras et al., 1992; Luong and Faugeras, 1992, 1994).

References

Ayache, N. 1990. *Stereovision and Sensor Fusion*. MIT Press.

Basu, A. 1993. Active calibration: Alternative strategy and analysis. In *Proc. of the Conf. on Computer Vision and Pattern Recognition*, New-York, pp. 495–500.

Brand, P., Mohr, R., and Bobet, P. 1993. Distorsions optiques: Correction dans un modele projectif. Technical Report RR-1933, INRIA.

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

Crowley, J., Bobet, P., and Schmid, C. 1993. Maintaining stereo calibration by tracking image points. In *Proc. of the Conf. on Computer Vision and Pattern Recognition*, New-York, pp. 483–488.

Deriche, R. and Blaszka, T. 1993. Recovering and characterizing image features using an efficient model based approach. In *Proc. International Conference on Computer Vision and Pattern Recognition*.

Deriche, R., Vaillant, R., and Faugeras, O. 1992. *From Noisy Edges Points to 3D Reconstruction of a Scene: A Robust Approach and Its Uncertainty Analysis*, Vol. 2, pp. 71–79. World Scientific. Series in Machine Perception and Artificial Intelligence.

Dron, L. 1993. Dynamic camera self-calibration from controled motion sequences. In *Proc. of the conf. on Computer Vision and Pattern Recognition*, New-York, pp. 501–506.

Du, F. and Brady, M. 1993. Self-calibration of the intrinsic parameters of cameras for active vision systems. In *Proc. of the conf. on Computer Vision and Pattern Recognition*, New-York, pp. 477–482.

Fang, J.Q. and Huang, T.S. 1984. 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.

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

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

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

Faugeras, O.D. and Maybank, S.J. 1990. Motion from point matches: Multiplicity of solutions. *The International Journal of Computer Vision*, 4(3):225–246, also INRIA Tech. Report 1157.

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

Garner, L.E. 1981. *An Outline of Projective Geometry*. Elsevier: North Holland.

Golub, G.H. and Van Loan, C.F. 1989. *Matrix Computations*. The John Hopkins University Press.

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

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

Hartley, R.I. 1994b. Self-calibration from multiple views with a rotating camera. In *Proc. European Conference on Computer Vision*, Stockholm, Sweden, pp. 471–478.

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

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

Jazwinsky, A.M. 1970. *Stochastic Processes and Filtering Theory*. Academic Press: London.

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

Kanatani, K. 1992. *Geometric Computation for Machine Vision*. Oxford university press.

Kruppa, E. 1913. Zur Ermittlung eines Objektes aus zwei Perspektiven mit innerer Orientierung. *Sitz.-Ber. Akad. Wiss., Wien, math. naturw. Kl., Abt. IIa.*, 122:1939–1948.

Kumar, R. and Hanson, A. 1990. Sensibility of the pose refinement problem to accurate estimation of camera parameters. In *Proceedings of the International Conference on Computer Vision*, Osaka, Japan, pp. 365–369.

Kumar, R.V.R., Tirumalai, A., and Jain, R.C. 1989. A non-linear optimization algorithm for the estimation of structure and motion parameters. In *Proc. International Conference on Computer Vision and Pattern Recognition*, pp. 136–143.

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

Luong, Q.-T. 1992. Matrice fondamentale et auto-calibration en vision par ordinateur. Ph.D. thesis, Universite de Paris-Sud, Orsay.

Luong, Q.-T. and Faugeras, O.D. 1992. Self-calibration of a camera using multiples images. In *Proc. International Conference on Pattern Recognition*, Den Hag, The Netherlands, pp. 9–12.

Luong, Q.-T. and Viéville, T. 1996. Canonic representations for the geometries of multiple projective views. *Computer Vision and Image Understanding*, 64(2):193–229.

Luong, Q.-T. and Faugeras, O.D. 1994. An optimization framework for efficient self-calibration and motion determination. In *Proc. International Conference on Pattern Recognition*, Jerusalem, Israel, pp. A-248–A-252.

Luong, Q.-T. and Faugeras, O.D. 1994. A stability analysis of the fundamental matrix. In *Proc. European Conference on Computer Vision*, Stockholm, Sweden, pp. 577–588.

Luong, Q.-T. and Faugeras, O.D. 1996. The fundamental matrix: Theory, algorithms, and stability analysis. *Intl. Journal of Computer Vision* 7(1):43–76.

Luong, Q.-T., Deriche, R., Faugeras, O.D., and Papadopoulo, T. 1993. On determining the fundamental matrix: Analysis of different methods and experimental results. Technical Report RR-1894, INRIA.

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

Maybank, S.J. and Faugeras, O.D. 1992. A theory of self-calibration of a moving camera. *The International Journal of Computer Vision*, 8(2):123–151.

Maybeck, P.S. 1979. *Stochastic Models, Estimation and Control*. Academic Press: London.

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

Robert, L. 1993. Reconstruction de courbes et de surfaces par vision stéréoscopique. Applications a la robotique mobile. Ph.D. thesis, Ecole Polytechnique.

Semple, J.G. and Kneebone, G.T. 1979. *Algebraic Projective Geometry*. Clarendon Press: Oxford, 1952 (Reprinted).

Spetsakis, M.E. and Aloimonos, J. 1988. Optimal computing of structure from motion using point correspondances in two frames. In *Proc. International Conference on Computer Vision*, pp. 449–453.

Trivedi, H.P. 1988. Can multiple views make up for lack of camera registration? *Image and Vision Computing*, 6(1):29–32.

Tsai, R.Y. 1989. Synopsis of Recent Progress on Camera Calibration for 3D Machine Vision. In Oussama Khatib, John J. Craig, and Tomás Lozano-Pérez (Eds.), *The Robotics Review*. MIT Press: pp. 147–159.

Tsai, R.Y. and Huang, T.S. 1984. 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.

Ullman, S. 1979. *The Interpretation of Visual Motion*. MIT Press.

Vieville, T. 1994. Auto-calibration of visual sensor parameters on a robotic head. *Image and Vision Computing*, 12.

Wampler, C.W., Morgan, A.P., and Sommese, A.J. 1988. Numerical continuation methods for solving polynomial systems arising in kinematics. Technical Report GMR-6372, General Motors Research Labs.

Weng, J., Ahuja, N., and Huang, T.S. 1989. Optimal motion and structure estimation. In *Proc. International Conference on Computer Vision and Pattern Recognition*, pp. 144–152.

Zhang, Z. and Faugeras, O.D. 1992. *3D Dynamic Scene Analysis*. Springer-Verlag.

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