Camera Calibration, Scene Motion and Structure recovery from point correspondences and fundamental matrices

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

Abstract

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. An application of the method to a binocular or trinocular stereo rig is also considered. It is illustrated by several experiments with real images which conclude the paper.

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 modelled 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 [53]. 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 [30] 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 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, Huang, and Maybank [20, 14, 36]. This matrix must

*Present address: University of California, EECS Cory Hall 211-215, Berkeley CA 94720

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 [14].

Assuming that the points being matched are attached to a plane makes things even simpler as shown for example by the work of Tsai and Huang [50] or that of Faugeras and Lustman [13]

When matching lines, things are a bit more complicated since at least three views are necessary to estimate the 3-D motion. Surprisingly enough, nonlinear estimation algorithms were discovered first, for example by Liu and Huang, Faugeras, Lustman, Toscani, and Spetsakis and Aloimonos [29, 16, 46]. The reason for this is that the analog of the E-matrix for lines is tensor of order three which has not yet been analysed as thoroughly as the E-matrix (see [55] though). Nonetheless, linear estimation algorithms have been published by Liu and Huang [28].

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. We also assume that we are given point correspondences, therefore excluding the case of lines. 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 modelling 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 illustrated in section 4. As an application of these ideas, in section 5 we assume that we have two or three cameras rigidly moving (a stereo rig) instead of one and show how to calibrate the rig completely, i.e. how to compute the intrinsic parameters of each camera and the relative displacement between the two or three. Finally, in section 7, 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, the intrinsic and extrinsic parameters, and 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 figure 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:

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

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 figure 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, a rotation of angle $\pi/2 - \theta$ around o followed by a translation $[u_0, v_0]^T$. The coordinate system (o, u, v) is the cordinate 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 idiosyncracies of the acquisition system. The angle θ models possible deviation from orthogonality of the cells' arrangement on the retina or a possible misalignment of the retinal plane with respect to the optical axis. In practice, it is usually quite close to $\pi/2$. The parameters relating the two retinal coordinate systems do not depend on the pose of the camera and are called the *camera intrinsic parameters.*

This camera model is essentially linear and ignores nonlinear effects such as those caused by lens distorsions. We assume either that they are not significant (which we found in most of our experiments) or that they have been corrected by standard techniques.

Thus no nonlinear camera distortion is considered which allows us to use the powerful tools of projective geometry. Projective geometry is emerging as an attractive framework for computer vision [39]. 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 [43, 5, 17],

but also in the computer vision litterature where it is presented in chapters of recent books [11, 23, 39], and articles [37, 22].

Using equation (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

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

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

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

Note that it 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 $\theta = \pi/2$, equation (3) takes the simpler form:

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

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

$$\mathbf{D} = \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}_3^T & 1 \end{pmatrix} \tag{5}$$

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 [14] 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 [37] 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

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

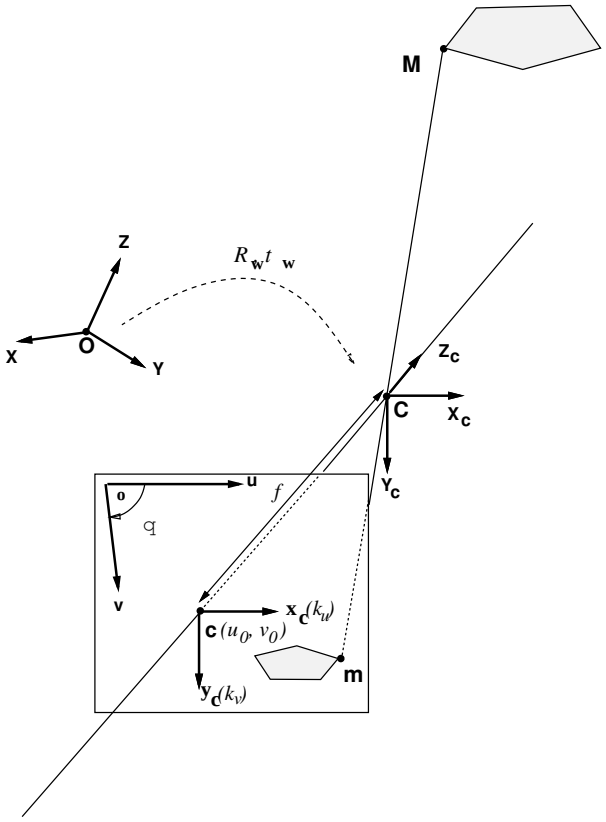

Figure 1: The general projective camera model

All points on that conic have complex coordinates. In fact, if we define $x = \frac{\mathcal{X}}{\mathcal{Z}}$ anf $y = \frac{\mathcal{Y}}{\mathcal{Z}}$, 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 know to Caylay. The proof of this can be found in [14, 11]. 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. 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. In fact, it is not difficult to show, and this is done in [11, 31], that the matrix defining the equation of the image of the absolute conic in the retinal coordinate system (o, u, v) is:

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

One of the important ideas which has emerged from our previous work [14, 37, 12] 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 figure 2.

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 $\mathbf{l}'_m$. If the relative camera geometry is known then, given a pixel m, its epipolar line l_m can be computed, and its correspondent m' has only to be searched along l_m rather than in the whole image.

Let us enrich this idea and consider the one parameter family of planes going through $\langle \mathbf{C}, \mathbf{C}' \rangle$ as shown in figure 3. This family is a pencil of planes. Let Π be any plane in the pencil, i.e. containing $\langle \mathbf{C}, \mathbf{C}' \rangle$. Then Π projects to an epipolar line l in the first image and to an epipolar line l' in the second image. The correspondences $\Pi \overline{\wedge} l$ and $\Pi \overline{\wedge} l'$ are homographies[1] between the two pencils of epipolar lines and the pencil of planes containing $\langle \mathbf{C}, \mathbf{C}' \rangle$. It follows that the correspondance $l \overline{\wedge} l'$ is a homography, called the epipolar transformation.

Now, in order to obtain an operational version of these properties, we are going to introduce an algebraic formulation, thanks to the key notion of *fundamental matrix*. It can be shown 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:

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

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

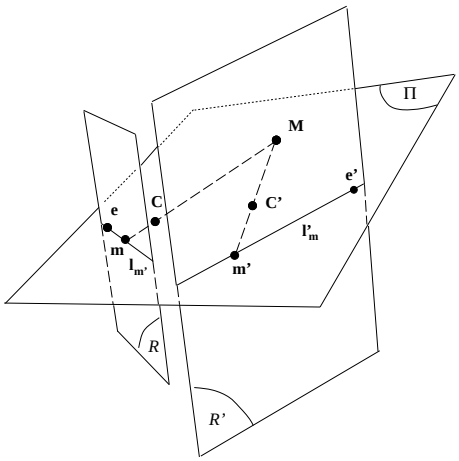

Figure 2: The epipolar geometry.

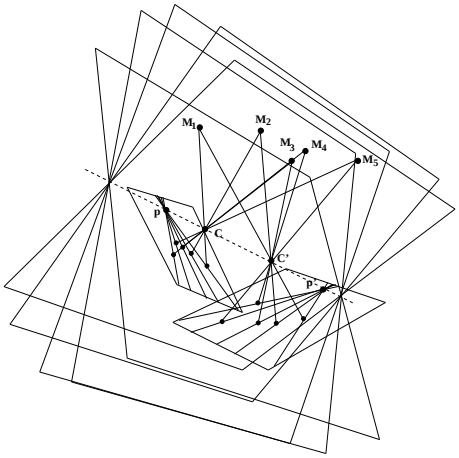

Figure 3: The epipolar pencils.

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.

In the practical case where epipoles are at finite distance, the epipolar transformation is characterized by the *affine* coordinates of the epipoles $\mathbf{e} = [e_1, e_2]^T$ and $\mathbf{e}' = [e'_1, e'_2]^T$ and by the coefficients of the homography between the two pencils of epipolar lines, each line being parameterized by its direction:

$$\tau \mapsto \tau^{'} = \frac{a\tau + b}{c\tau + d} \tag{9}$$

where

$$\tau = \frac{m_2 - e_2}{m_1 - e_1} \qquad \tau' = \frac{m_2^{'} - e_2^{'}}{m_1^{'} - e_1^{'}} \tag{10}$$

and $\mathbf{m} \leftrightarrow \mathbf{m}'$, is a pair of corresponding points. It follows that the epipolar transformation, like the fundamental matrix depends on seven independent parameters.

Equation (8) is the analog in the uncalibrated case of the so-called Longuet-Higgins equation [30]. 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 figure 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:

$$\mathbf{m}' \cdot (\mathbf{t} \times \mathbf{Rm}) \equiv \mathbf{m}^{'T}\mathbf{Em} = 0 \tag{11}$$

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 equations (11) and (8) are equivalent, and that we have the relation:

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

Unlike the essential matrix, which is characterized by the two constraints found by Huang and Faugeras [20] which are the nullity of the determinant and the equality of the two non-zero singular values, the only property of the fundamental matrix is that it is of rank two. As it is also defined only up to a scale factor, the number of independent coefficients of $\mathbf{F}$ is seven, as seen previously.

2.3 The rigidity constraint, Kruppa equations and the intrinsic parameters

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. 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 independant 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:

$$\mathbf{E} = \mathbf{A}^{'T}\mathbf{FA} \tag{12}$$

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 (12) is an essential matrix :

$$\det(\mathbf{E}) = 0 \qquad f(\mathbf{E}) = \frac{1}{2}\text{trace}^2(\mathbf{E}\mathbf{E}^T) - \text{trace}(\mathbf{E}\mathbf{E}^T)^2 = 0 \tag{13}$$

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 [48]. If $\mathbf{E}$ is an essential matrix, the symmetric matrix $\mathbf{S} = \mathbf{E}\mathbf{E}^T$, which *a priori* has six independent entries, depends only on the three components of $\mathbf{t}$:

$$\mathbf{E}\mathbf{E}^T = -[\mathbf{t}]_\times{}^2 = \begin{bmatrix} t_2^2 + t_3^2 & -t_1 t_2 & -t_1 t_3 \\ -t_2 t_1 & t_3^2 + t_1^2 & -t_2 t_3 \\ -t_3 t_1 & -t_3 t_2 & t_1^2 + t_2^2 \end{bmatrix} \tag{14}$$

The matrix $\mathbf{S} = \mathbf{E}\mathbf{E}^T$ 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}), $1 \leq i < j \leq 3$:

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

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.1 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 [24] 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 [14], and then to develop a theory of self-calibration [37]. In this exposition, we will return to the original formulation, which doesn't assume that the two cameras are identical.

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 figure 4.

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 is the dual matrix (matrix of cofactors) of $\mathbf{B}$:

$$\mathbf{K} = \mathbf{B}^* \tag{15}$$

whose entries are given using the notations of Kruppa, and called Kruppa coefficients:

$$\mathbf{K} = \begin{pmatrix} -\delta_{23} & \delta_3 & \delta_2 \\ \delta_3 & -\delta_{13} & \delta_1 \\ \delta_2 & \delta_1 & -\delta_{12} \end{pmatrix} \tag{16}$$

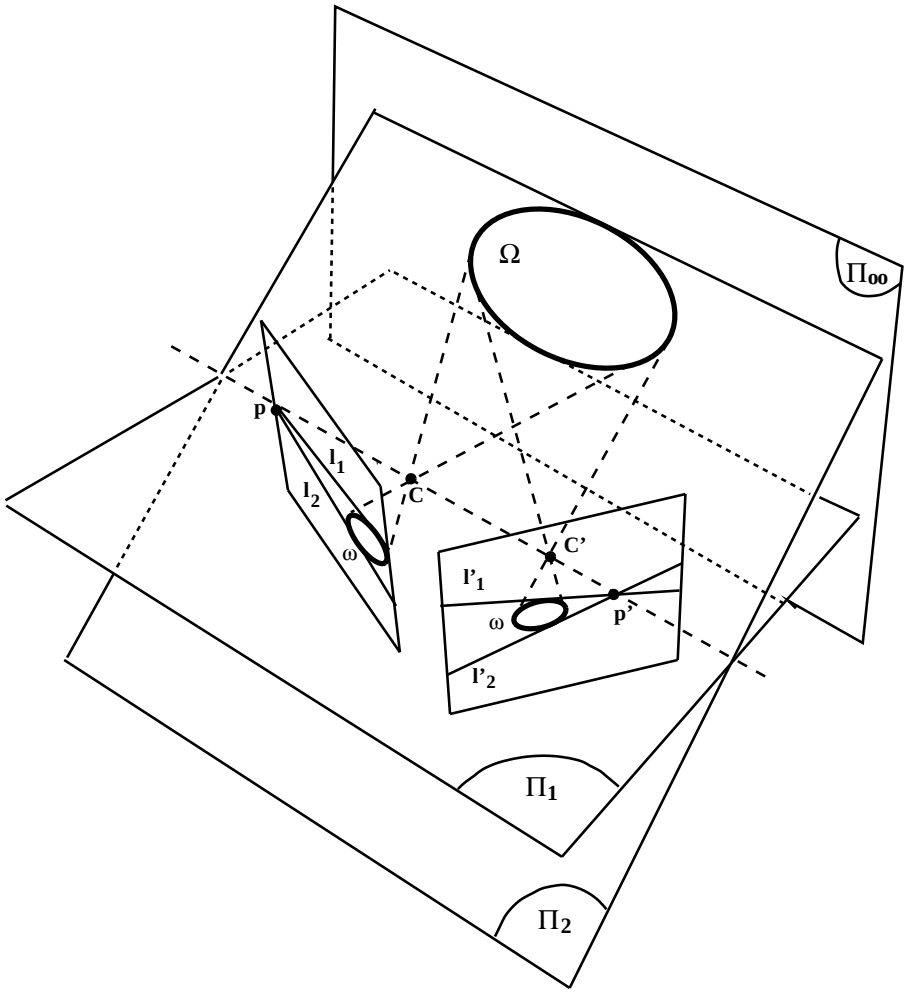

Figure 4: The absolute conic and epipolar transformation

Thus by definition, we have the relations:

$$\begin{aligned}\delta_1 &= b_{12}b_{13} - b_{11}b_{23}\\ \delta_2 &= b_{23}b_{21} - b_{22}b_{31}\\ \delta_3 &= b_{31}b_{32} - b_{33}b_{12}\end{aligned} \tag{17}$$

and

$$\begin{aligned}\delta_{12} &= b_{12}^2 - b_{11}b_{22}\\ \delta_{23} &= b_{23}^2 - b_{22}b_{33}\\ \delta_{31} &= b_{31}^2 - b_{33}b_{11}\end{aligned} \tag{18}$$

In the second image, we have the same relations, which define a matrix $\mathbf{K}' = \mathbf{B}'^{*}$, by its entries δ_i', δ_{ij}'.

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

$$(\mathbf{e} \times \mathbf{y})^\top \mathbf{K} (\mathbf{e} \times \mathbf{y}) = 0$$

by parameterizing the epipolar line l with the projective parameter τ such that $\mathbf{y} = (1, \tau, 0)^T$, this equation can be written[2]:

$$P_1(\tau) = k_{11} + 2k_{12}\tau + k_{22}\tau^2 = 0 \tag{19}$$

where the coefficients k_{11}, k_{12}, k_{22} are defined by

$$\begin{aligned}k_{11} &= -\delta_{13} - \delta_{12}e_2^2 - 2\delta_1 e_2\\ k_{12} &= \delta_{12}e_1e_2 - \delta_3 + \delta_2 e_2 + \delta_1 e_1\\ k_{22} &= -\delta_{23} - \delta_{12}e_1^2 - 2\delta_2 e_1\end{aligned} \tag{20}$$

Similarly, the epipolar line l' through $\mathbf{e}'$ corresponding to l is tangent to ω':

$$k_{11}^{'} + 2k_{12}^{'}\tau' + k_{22}^{'}\tau'^2 = 0 \tag{21}$$

Since l' correspond to l, its projective parameter τ' is obtained from the projective parameter τ by (9), and thus (21) can be written, after substitution:

$$P_2(\tau) = k_{11}'' + 2k_{12}''\tau + k_{22}''\tau^2 = 0 \tag{22}$$

with:

$$\begin{aligned}k_{11}^{''} &= k_{22}'b^2 + k_{11}'d^2 + 2k_{12}'bd;\\ k_{12}^{''} &= 2k_{12}'ad + 2k_{22}'ab + 2k_{11}'cd + 2k_{12}'bc\\ k_{22}^{''} &= 2k_{12}'ac + k_{22}'a^2 + k_{11}'c^2\end{aligned}$$

the coefficients $k_{11}^{'}$, $k_{12}^{'}$, $k_{22}^{'}$ being obtained from (20) by replacing the coordinates e_i of $\mathbf{e}$ with the coordinates $e_i^{'}$ of $\mathbf{e}^{'}$ and the coefficients δ_i, δ_{ij} with the coefficients δ_i', δ_{ij}'.

The polynomials P_1 and P_2 must have the same roots, which yield three so-called Kruppa equations, of which only two are independant:

$$\begin{aligned}k_{22}k_{12}^{''} - k_{22}^{''}k_{12} &= 0\\ k_{11}k_{12}^{''} - k_{11}^{''}k_{12} &= 0\\ k_{11}k_{22}^{''} - k_{11}^{''}k_{22} &= 0\end{aligned} \tag{23}$$

[2]This assumes that the epipole e is not at infinity. In that case we write $\mathbf{e} = [e_1, e_2, 1]^T$.

It can be 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 equations (23) is equivalent to the second constraint, which is done in [31, 34].

The nice thing with the Kruppa equations is that they are only of degree two in the twelve parameters δ_i, δ_{ij}, δ'_i, δ'_{ij}, 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

3.1 Using three images taken by a moving camera

A moving camera In an earlier work, Trivedi [48] 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}) mentionned 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 [18] 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:

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

He exhibits an algorithm to factor the fundamental matrix $\mathbf{F}$ as $\mathbf{A}'^{-1T}\mathbf{E}\mathbf{A}^{-1}$, 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 additionnal equations. The idea behind our method is to use constraints which arise from the observation of a static scene by a *single*[3] moving camera. In this case the intrinsic parameters remain constant: $\mathbf{A} = \mathbf{A}'$, $\delta_i = \delta'_i$, $\delta_{ij} = \delta'_{ij}$, thus we can cumulate constraints over different displacements, and obtain a sufficient number of equations for the resolution.

How many displacements are necessary ? Each displacement yields two independent algebraic equations. In the case of a moving camera, we have only five coefficients δ_i, δ_{ij}, to estimate, since they are 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 (33).

[3] 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 than with multiple different cameras.

But if we perform two displacements, we will obtain three images, 1,2,3. Between three images, there are in fact 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 underconstrained system. One way to see that it is not the case is to note that in our case where we consider displacements only up to a scale factor, the translational part of displacement $\mathbf{D}_3$ *cannot* be obtained from displacements $\mathbf{D}_1$ and displacement $\mathbf{D}_2$, since the norm of the translations is unknown, as pointed out in section 5.1. Let us rephrase the argument in another way. Two fundamental matrices depend upon 14 parameters. But if we are to achieve self-calibration, then we have eventually to describe the three displacements 1-2, 2-3, and 1-3 up to a scale factor (11 parameters, as shown in section 5.1) and the 5 intrinsic parameters. This is a function of 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 additionnal Kruppa equations provided by the third fundamental matrix. We also give in appendix B a simple numerical exemple 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, it is known [47, 49] that a simple method works well in this case.

In the more frequent case where the displacement is a pure translation $\mathbf{t} = (t_1, t_2, t_3)^T$, the rotation is the identity $\mathbf{R} = \mathbf{I}_3$. The fundamental matrix being antisymetric, it is easy to see that the epipoles $\mathbf{e}$ and $\mathbf{e}'$ are the same, and the homography is the identity $\tau \mapsto \tau$, resulting in Kruppa equations which reduce to tautologies. The coefficients k'_{ij} from (23) and the coefficients k_{ij} from (20) are identical, since the epipoles are the same, the coefficients k''_{ij} and k'_{ij} from (23) are identical, since the homography is the identity. Thus the two polynomials P_1 and P_2 are equal. A geometric interpretation is that since the two tangents to ω in the first image are the same as the tangents to ω in the second image, and the epipolar transformation is the identity, no constraint can be derived from the rigidity.

Intrinsic parameters and Kruppa coefficients We now show that in the case of a unique moving camera there is a one-to-one correspondence between the Kruppa coefficients previously introduced, and the real intrinsic parameters.

Using the relations (15) and (7), we obtain the relation: $\mathbf{K} = \mathbf{A}\mathbf{A}^T$, defined up to a scale factor λ. Using the definition (16) of $\mathbf{K}$, yields:

$$\lambda\delta_1 = v_0 \tag{24}$$

$$\lambda\delta_2 = u_0 \tag{25}$$

$$\lambda\delta_3 = u_0 v_0 - \alpha_u\alpha_v \frac{\cot\theta}{\sin\theta} \tag{26}$$

$$\lambda\delta_{12} = -1 \tag{27}$$

$$\lambda\delta_{23} = -u_0^2 - \frac{\alpha_u^2}{\sin^2\theta} \tag{28}$$

$$\lambda\delta_{13} = -v_0^2 - \frac{\alpha_v^2}{\sin^2\theta} \tag{29}$$

By combining (25), (28), and (27):

$$\alpha_u^2 = \frac{\delta_{23}\delta_{12} - \delta_2^2}{\delta_{12}^2}\sin^2\theta$$

and similarly, with (24), (29), et (27):

$$\alpha_v^2 = \frac{\delta_{13}\delta_{12} - \delta_1^2}{\delta_{12}^2} \sin^2\theta$$

By substitution of these two relations in (26), we obtain:

$$\cos^2\theta = \frac{(\delta_3\delta_{12} + \delta_1\delta_2)^2}{(\delta_{13}\delta_{12} - \delta_1^2)(\delta_{23}\delta_{12} - \delta_2^2)}$$

Thus we can conclude that the system of equations (24-29) has a real solution if and only if the following three conditions are satisfied:

$$\begin{cases} \delta_{13}\delta_{12} - \delta_1^2 > 0 \\ \delta_{23}\delta_{12} - \delta_2^2 > 0 \\ (\delta_{13}\delta_{12} - \delta_1^2)(\delta_{23}\delta_{12} - \delta_2^2) \geq (\delta_3\delta_{12} + \delta_1\delta_2)^2 \end{cases} \tag{30}$$

Among these conditions, only two are independent, specifically if the third condition is satisfied, then the first two are equivalent. I has be shown in [31, 34] that these conditions are equivalent to the fact that the conic ω whose matrix is $\mathbf{K}^*$ has no real point.

Some algebra then yields the intrinsic parameters:

$$\begin{aligned} u_0 &= -\frac{\delta_2}{\delta_{12}} \\ v_0 &= -\frac{\delta_1}{\delta_{12}} \\ \alpha_u &= \varepsilon\sqrt{\frac{\delta_{23}\delta_1^2 + \delta_{13}\delta_2^2 + \delta_{12}\delta_3^2 + 2\delta_1\delta_2\delta_3 - \delta_{12}\delta_{23}\delta_{13}}{\delta_{12}(\delta_1^2 - \delta_{13}\delta_{12})}} \\ \alpha_v &= \varepsilon\sqrt{\frac{\delta_{23}\delta_1^2 + \delta_{13}\delta_2^2 + \delta_{12}\delta_3^2 + 2\delta_1\delta_2\delta_3 - \delta_{12}\delta_{23}\delta_{13}}{\delta_{12}(\delta_2^2 - \delta_{23}\delta_{12})}} \\ \cos\theta &= \frac{\delta_3\delta_{12} + \delta_2\delta_1}{\sqrt{(\delta_{23}\delta_{12} - \delta_2^2)(\delta_{13}\delta_{12} - \delta_1^2)}} \end{aligned} \tag{31}$$

$$\varepsilon = \pm 1 \tag{32}$$

These expression show that they are uniquely determined, except for the sign of the scale factors α_u et α_v. From these relations, two remarks can be made. The first one is that the intrinsic parameters depend only on the ratio of the Kruppa coefficients, which was expected. The second one is that it is quite simple to formulate the four-parameter model. Since in that case $\theta = \frac{\pi}{2}$, the third constraint (30) becomes simply:

$$\delta_3\delta_{12} + \delta_2\delta_1 = 0 \tag{33}$$

3.2 A semi-analytic method

Principle Three displacements yield six equations in the entries of the matrix $\mathbf{K}$ defined in (16). 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 prefered 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 [56] 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 coefficients,
  - keep only the real solutions $\mathcal{K}_i^k$ which satisfy the constraints (30),
- for each pair of solution lists $\{\mathcal{K}_i^k\}_k$, $\{\mathcal{K}_j^k\}_k$, increment a counter corresponding to the solution in the list i and the one in the list j which minimize the distance[4]:

$$d(\mathbf{u}, \mathbf{v}) = \sum_{i=1}^{5} \frac{\|u_i - v_i\|}{\max(\|u_i\|, \|v_i\|)}$$

  where $\mathbf{u}$ and $\mathbf{v}$ are two 5-dimensional vectors representing individual solutions, and u_i and v_i are their components,
- pick up in each list $\mathcal{K}_i$ the solution which has obtained the highest counter score,
- compute the intrinsic parameters using the formulas (31),
- compute the final solution and an estimate of the covariance by an averaging operator[5].

Two examples We have also verified the impact of the orthogonality constraint (33) 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 (33), which is

[4] This relative distance had to be chosen, because the orders of magnitude of each component are very different.

[5] 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.

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 (33). However, for the sake of comparison, we have used only six of these equations.

We present in table 1 and table 2 two typical examples[6] of results obtained by each method, using synthetic data which are point correspondences at different noise levels. 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 (30) was found),
- 3 displacements with the orthogonality constraint,
- 3 displacements without the orthogonality constraint.

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 [32] from these point correspondences, and then use the continuation algorithm to solve the Kruppa equations obtained from the fundamental matrices. Numbers in brackets are estimates of the uncertainty of the results.

noise	method		Estimated parameters				
(pixels)	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	10^{-13}
	Y	2,3	639.44	944.36	246.04	258.59	10^{-13}
	Y	1,3	641.62	945.73	248.97	255.56	10^{-12}
	Y	1,2,3	641.69 [2.0]	947.49 [3.7]	247.03 [1.2]	256.55 [1.7]	10^{-13} [10^{-12}]
	N	1,2,3	644.40 [2.3]	952.29 [3.8]	237.45 [4.2]	254.61 [1.9]	6.10^{-3} [10^{-3}]
0.5	Y	1,2	651.39	962.47	244.73	246.56	10^{-12}
	Y	2,3	636.54	946.72	245.82	270.46	10^{-12}
	Y	1,3	647.41	953.67	260.91	255.11	10^{-13}
	Y	1,2,3	648.39 [11.1]	963.69 [20.3]	250.84 [6.7]	260.26 [9.1]	10^{-13} [10^{-12}]
	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	10^{-13}
	Y	1,3	74.85	455.95	733.93	434.07	10^{-10}
	Y	1,2,3	658.00 [24.8]	986.63 [45.7]	255.61 [14.3]	265.09 [19.7]	10^{13} [10^{-12}]
	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	10^{-12}
	Y	2,3	627.92	950.11	245.16	300.56	10^{-12}
	Y	1,3	659.79	971.19	293.82	252.73	10^{-13}
	Y	1,2,3	669.62 [42.6]	1013.85 [79.2]	260.16 [23.3]	270.23 [32.3]	10^{-12} [10^{-12}]
	N	1,2,3	633.02 [73.0]	1223.62 [104.5]	190.46 [231.1]	205.49 [43.9]	0.27 [0.2]

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

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:

- 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,

[6] It can be seen that results can differ significantly from one configuration to another. For a statistical approach and a more global assessment of the precision, see next section.

noise (pixels)	method orth.	method displ.	estimated parameters 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	10^{-13}
	Y	2,3	124.317	947.934	230.705	252.053	10^{-13}
	Y	1,3	639.303	943.591	246.083	257.594	10^{-12}
	Y	1,2,3	640.83 [2.7]	947.59 [2.7]	237.90 [10.2]	252.88 [3.1]	10^{-13} $[10^{-12}]$
	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	10^{-12}
	Y	1,2,3	654.01 [24.1]	976.83 [22.4]	214.28 [47.0]	232.66 [20.7]	10^{-13} $[10^{-12}]$
	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	10^{-13}
	Y	2,3	-	-	-	-	-
	Y	1,3	630.86	941.71	245.96	277.04	10^{-13}
	Y	1,2,3	505.94 [248.7]	779.03 [389.9]	179.30 [94.4]	407.68 [317.3]	10^{13} $[10^{-12}]$
	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	10^{-14}
	Y	2,3	342.86	875.35	219.38	246.91	10^{-14}
	Y	1,3	604.46	885.15	249.27	260.23	10^{-12}
	Y	1,2,3	688.43 [163.9]	1048.77 [254.3]	161.48 [75.0]	207.08 [38.7]	10^{-12} $[10^{-12}]$
	N	1,2,3	1190.91 [1164.0]	1803.80 [1867.3]	109.39 [149.1]	-109.65 [661.5]	0.13 [0.1]

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

- 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 (30) 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 iterative formulation.

3.3 Iterative formulations

In this approach, we do no longer make use of the simple polynomial structure of the Kruppa equations, but rather consider them as measurement equations relating directely fundamental matrices to intrinsic parameters, obtained by substituting the values (24-29) into (23). We can then solve them either by a batch non-linear least-squares minimization technique, or by an extended Kalman filtering approach.

Global minimization The choice of the criterion to be minimized is very important. We have noticed two things. First, using the three Kruppa equations even though they are not independent provides additional constraints and improve the results. Second, minimizing directly the value of the residual of expressions (23) do not work well. The reason is the well known fact that minimizing a criterion $\sum_i (\frac{a_i}{b_i} - \frac{a'_i}{b'_i})^2$ is quite different from minimizing $\sum_i (a_i b'_i - a'_i b_i)^2$ 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 from (19) and (22) we thus try to minimize the following criterion:

$$\min_{\alpha_u,\alpha_v,u_0,v_0,\theta} \sum_i \left(\frac{k_{11}}{k''_{11}} - \frac{k_{12}}{k''_{12}}\right)^2 + \left(\frac{k_{12}}{k''_{22}} - \frac{k_{12}}{k''_{22}}\right)^2 + \left(\frac{k_{11}}{k''_{22}} - \frac{k_{11}}{k''_{22}}\right)^2 \tag{34}$$

where the coefficients k_{ij} and k''_{ij} are defined in (20) and (23).

We have compared this method with the continuation method, using a statistical approach involving 100 triples[7] 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 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.

noise (pixels)	method	failure	parameters				
			α_u	α_v	u_0	v_0	$\theta - \frac{\pi}{2}$
0.2	Continu	2	0.056	0.062	0.133	0.136	0.035
	Mini (1)		0.025	0.028	0.071	0.095	0.034
	Mini (2)		0.054	0.056	0.106	0.123	0.051
	Mini (3)		0.065	0.063	0.136	0.153	0.055
1.0	Continu	7	0.120	0.141	0.263	0.321	0.080
	Mini (1)		0.115	0.141	0.261	0.327	0.098
	Mini (2)		0.138	0.160	0.274	0.370	0.130
	Mini (3)		0.174	0.184	0.281	0.360	0.135

Table 3: Statistical results with 3 displacements (see text).

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,
- 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[8], 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 and f_2 are two Kruppa equations (23)

[7] The results improve if one considers more displacements. See next paragraph.

[8] 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 [21] [38], or the more practical presentations which can be found in [1], [11], and [59].

The perpendicularity 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. Figure 5 shows an example obtained with 0.5 pixel of image noise. The convergence happens between 5 and 10 displacements.

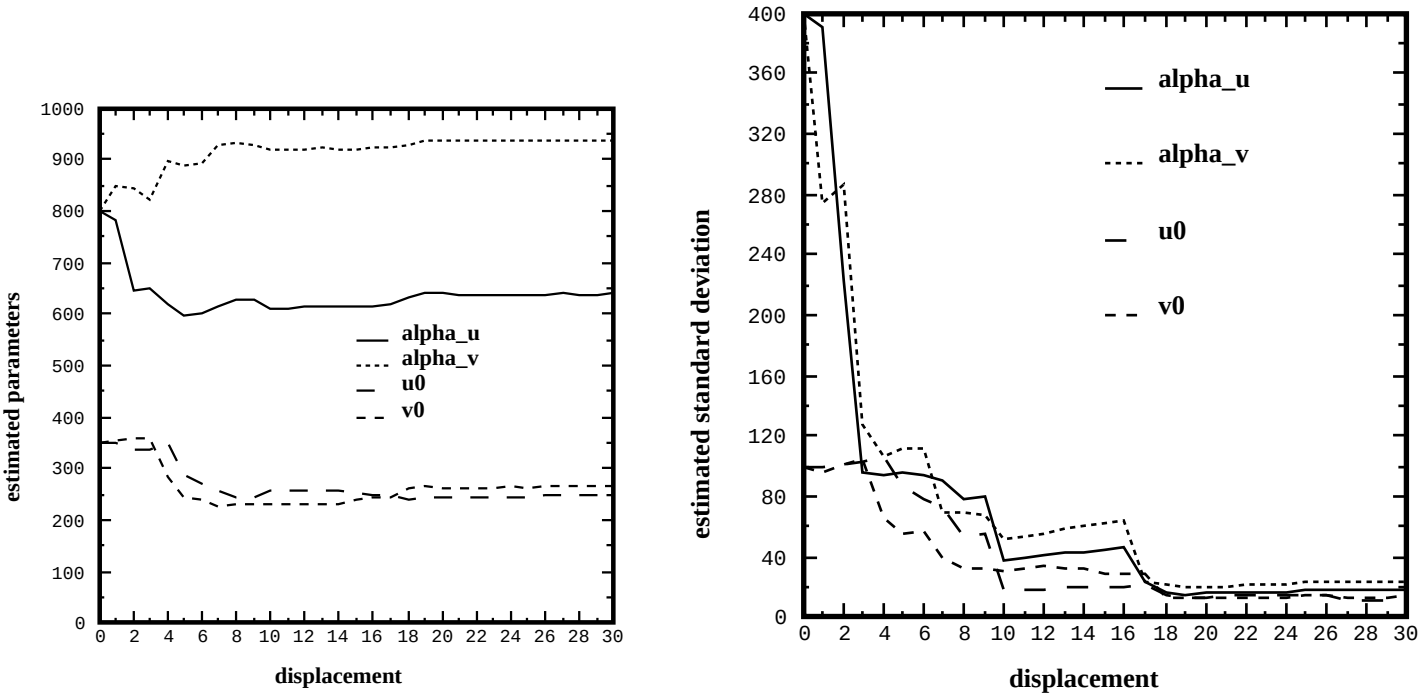

Figure 5: An example of computation of intrinsic parameters by Kalman filtering.

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[9]. 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 [57] and [26]. 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 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 method

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 the case of minimal number of displacements, the continuation method seems however preferable, whereas the iterative approaches are well suited to the case where more displacements are available.

[9] The apparingly less good 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

noise (pixel)	nb displ	α_u mini	α_u kalman	α_v mini	α_v kalman	u_0 mini	u_0 kalman	v_0 mini	v_0 kalman
0.2	3	0.1494	0.1389	0.1487	0.1398	0.3047	0.2778	0.3091	0.2840
	5	0.0952	0.1377	0.0837	0.1462	0.2264	0.2676	0.2591	0.2713
	10	0.0652	0.1390	0.0714	0.1333	0.1970	0.2343	0.2401	0.2407
	15	0.0464	0.1392	0.0753	0.1201	0.2033	0.2243	0.2042	0.2167
1.0	3	0.3886	0.3535	0.3944	0.3647	0.5237	0.5070	0.5783	0.5540
	5	0.3335	0.3383	0.3124	0.3664	0.4742	0.4936	0.499	0.5221
	10	0.2875	0.3534	0.2913	0.3707	0.4406	0.4644	0.5144	0.4888
	15	0.3017	0.3662	0.2710	0.3712	0.4336	0.4348	0.4954	0.4774

Table 4: Comparison of minimization and Kalman filtering (1).

noise (pixel)		α_u mini	α_u kalman	α_v mini	α_v kalman	u_0 mini	u_0 kalman	v_0 mini	v_0 kalman
0.2	err	0.0412	0.0408	0.0737	0.0659	0.1832	0.1448	0.1916	0.1396
	% err > 0.05	22	13	28	15	48	29	58	40
1.0	err	0.2337	0.2498	0.2245	0.2796	0.4165	0.3500	0.4854	0.4171
	% err > 0.05	68	49	64	58	78	59	78	72

Table 5: Comparison of minimization and Kalman filtering (2).

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 [31, 33].

Another limitation might be that the method does not give an accurate estimation for the position of the principal point, and the angle of 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.

4 Taking into account the motion of the camera

We suppose now that we have obtained the intrinsic parameters $\mathbf{A}$ of a camera. Our next goal is to compute the three-dimensionnal motion from pairs of images. This computation can be done quite robustly even with imprecise camera parameters. We can take advantage of this remark to combine this computation with the computation of intrinsic parameters. We obtain another iterative approach to self-calibration, which yields more robust results than the Kruppa approach.

4.1 Computing the motion after calibrating

The motion determination problem from point correspondences is very classical. See [16] [45] [57] [19] for solutions similar to ours. We present 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 $\mathbf{F}$, from which the essential matrix is immediately obtained:

$$\mathbf{E} = \mathbf{A}^T \mathbf{F} \mathbf{A} \tag{35}$$

The problem of finding the rotation $\mathbf{R}$ and the translation $\mathbf{t}$ from $\mathbf{E}$ is classical [30, 51, 16].

As we have, by construction, found a F-matrix of rank two, the direction of translation is just obtained by solving: $\mathbf{E}^T \mathbf{t} = 0$.

To find the rotation, we use a method introduced by [16]: in the presence of noise, we minimize with respect to the rotation matrix $\mathbf{R}$ the criterion:

$$C = \sum_{i=1}^{3} \|\mathbf{E}_i - \mathbf{R}^T \mathbf{T}_i\|^2$$

where $\mathbf{E}_i$ and $\mathbf{T}_i$ are the three rows of the matrices $\mathbf{E}$ and $\mathbf{T}$, respectively. Using $\mathbf{q}$ a quaternion representing $\mathbf{R}$, some properties of this representation yield:

$$C = \sum_{i=1}^{3} \mid \mathbf{q} \times \mathbf{E}_i - \mathbf{T}_i \times \mathbf{q} \mid^2 \tag{36}$$

where $\times$ denotes the quaternion product. It follows from the definition of the quaternion product that $\mathbf{q} \times \mathbf{E}_i - \mathbf{T}_i \times \mathbf{q}$ is a linear function of the 4 coordinates of $\mathbf{q}$. Therefore, there exists a 4×4 matrix $\mathbf{N}_i$ such that:

$$\mid \mathbf{q} \times \mathbf{E}_i - \mathbf{T}_i \times \mathbf{q} \mid = \mathbf{N}_i \mathbf{q} \quad \text{with} \quad \mathbf{N}_i = \begin{pmatrix} 0 & (\mathbf{E}_i - \mathbf{T}_i)^T \\ \mathbf{T}_i - \mathbf{E}_i & [\mathbf{E}_i]_\times + [\mathbf{T}_i]_\times \end{pmatrix} \tag{37}$$

Therefore, the problem reduces to a linear least-squares problem:

$$\min_{\mathbf{q}} \sum_{i=0}^{3} \mathbf{q} \mathbf{N}_i \mathbf{N}_i^T \mathbf{q}^T \quad \text{subject to the constraint:} \quad \|\mathbf{q}\|^2 = 1$$

which is a classical minimization problem, whose solution is the eigenvector associated with the smallest eigenvalue of $\mathbf{N} = \sum_{i=1}^{3} \mathbf{N}_i \mathbf{N}_i^T$. It can be noted that this solution is entirely equivalent to the well-known method of Tsai and Huang [51] , which has been recently proved to be optimal by Hartley [18] . We denote this algorithm by **FACTOR**.

An iterative solution An alternative method is to use directly the criterion that has been used to determine the fundamental matrix. In [32] different parametrizations 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 then clearly show that the linear criterion:

$$\min_{\mathbf{F}} \sum_{i} (\mathbf{m}_i'^T \mathbf{F} \mathbf{m}_i)^2 \quad \text{subject to } \mathrm{Tr}(\mathbf{F}^T \mathbf{F}) = 1 \tag{38}$$

is unable to express the rank and normalization constraints. Using the linear criterion leads definitely to the worst result in the determination of the fundamental matrix. To overcome the major weaknesses of the linear criterion, different non-linear criteria were proposed and analyzed in great detail. We have found that the following criterion works well:

$$\min_{\mathbf{F}} \{ d(\mathbf{m}'^T, \mathbf{F}\mathbf{m})^2 + d(\mathbf{m}^T, \mathbf{F}^T \mathbf{m}')^2 \} \tag{39}$$

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

We denote by **MIN-LIN**[10] the minimization of the error criterion (38) and by **MIN-DIST** the minimization of the error criterion (39). The knowledge of the intrinsic parameters allows us to minimize these criteria with respect to five motion parameters: we parameterize $\mathbf{T}$ by t_1/t_3, t_2/t_3 and $\mathbf{R}$ by the three-dimensional vector $\mathbf{r}$ whose direction is that of the axis of rotation and whose norm is equal to the rotation angle. We use, as a starting point for this non-linear minimization, the result obtained by **FACTOR**.

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. As 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 (figures 6 and 7).
2. the motion obtained by **FACTOR**, which is the realistic initialization (figures 8 and 9, 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 (figures 8 and 9, label 2).

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.
- 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 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 [51, 10].

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 [25]. 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 dimensionnal reconstruction, which is the final objective. We present here some experimental results which give an idea of the numerical values. Figure 10 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 11 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).

[10] 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.

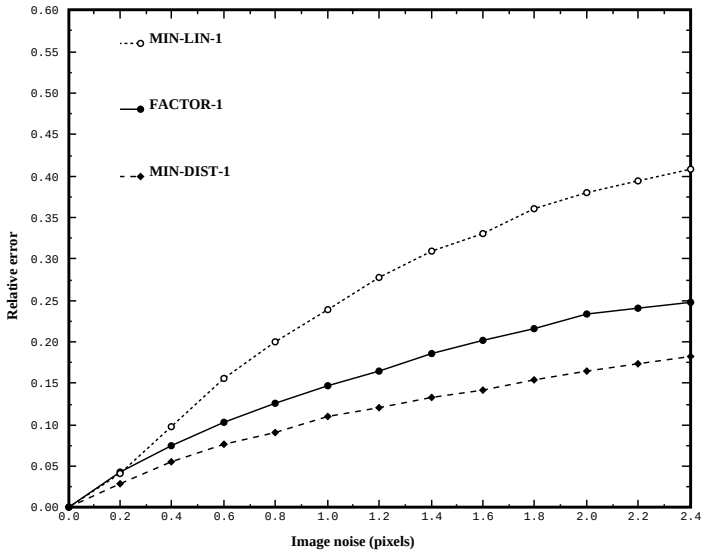

Figure 6: Relative error on the rotation, initialization with the exact displacement.

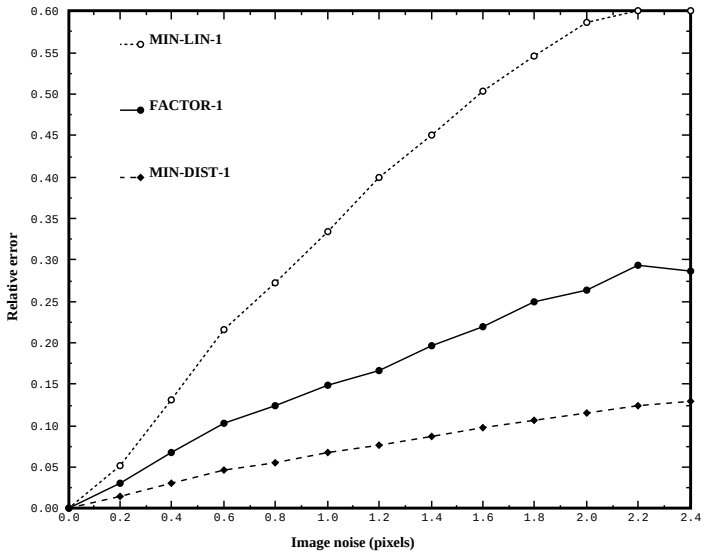

Figure 7: Relative error on the translation, initialization with the exact displacement.

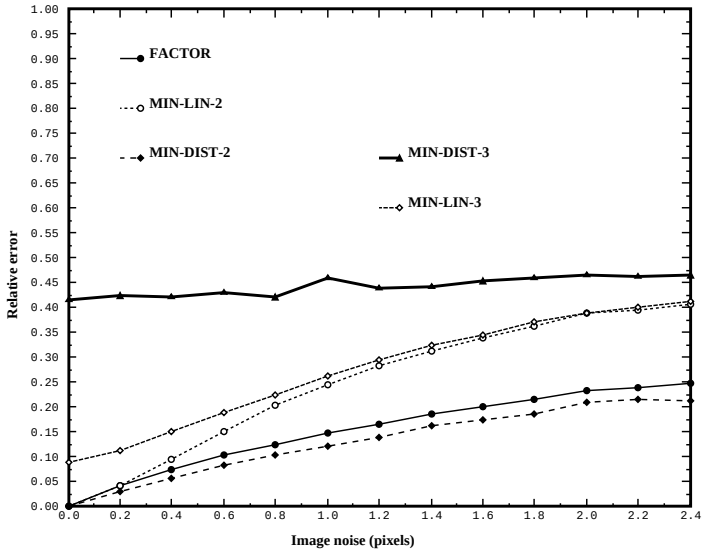

Figure 8: Relative error on the rotation, initialization with two different values (see text).

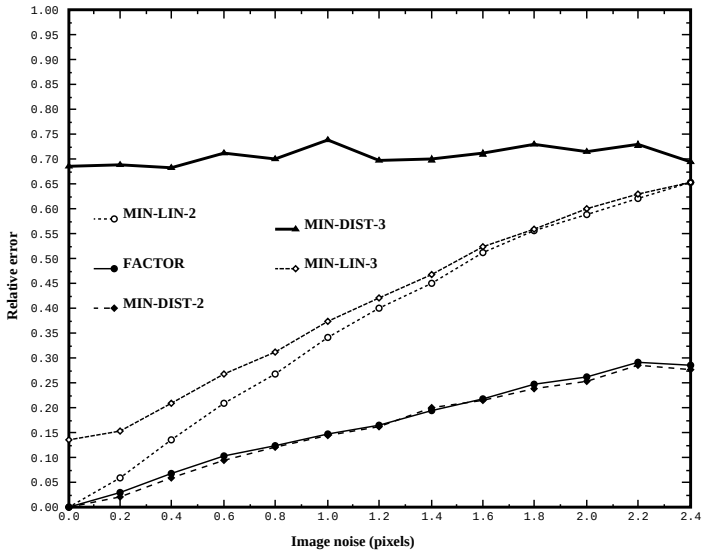

Figure 9: Relative error on the translation, initialization with two different values (see text).

- 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 is 50% error and has the same effects as a 25% 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 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.

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 criterion (39) simultaneously with respect to the five motion parameters previously introduced and to the intrinsic parameters, by substitution of $\mathbf{A}^{-1T}\mathbf{T}\mathbf{R}\mathbf{A}^{-1}$ in place of $\mathbf{F}$. Although the number of equations appearing in the least-squares formulation is the number of point matches, which may be very large, we must remember from the analysis made in section 2.3 that only two parameters can be computed. Since we have seen that the most significant are α_u and α_v, we chose to allow them to vary and to leave u_0 and v_0 fixed. The relative errors obtained on the motion parameters are shown in figure 12. They are to be compared to figure 11, and to facilitate this comparison, we have also plotted on this figure the two curves obtained in figure 11 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 figure 13, which shows that the final error on the camera parameters depends mainly on the noise, and not on the initial error on the parameters until 25%, where some convergence problems appear. They are revealed by the fact that the final error increases whereas the noise level remains constant. If the algorithm achieved perfect convergence, the same solution could be found for a given noise level, and thus the final error would not depend at all from the starting value.

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 to further constrain the problem, if more than three displacements are available. 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 accomodate N independant displacements ($N \geq 2$), and, for each displacement i, a minimum of eight correspondences $(\mathbf{q}_{ij}, \mathbf{q}'_{ij})_j$:

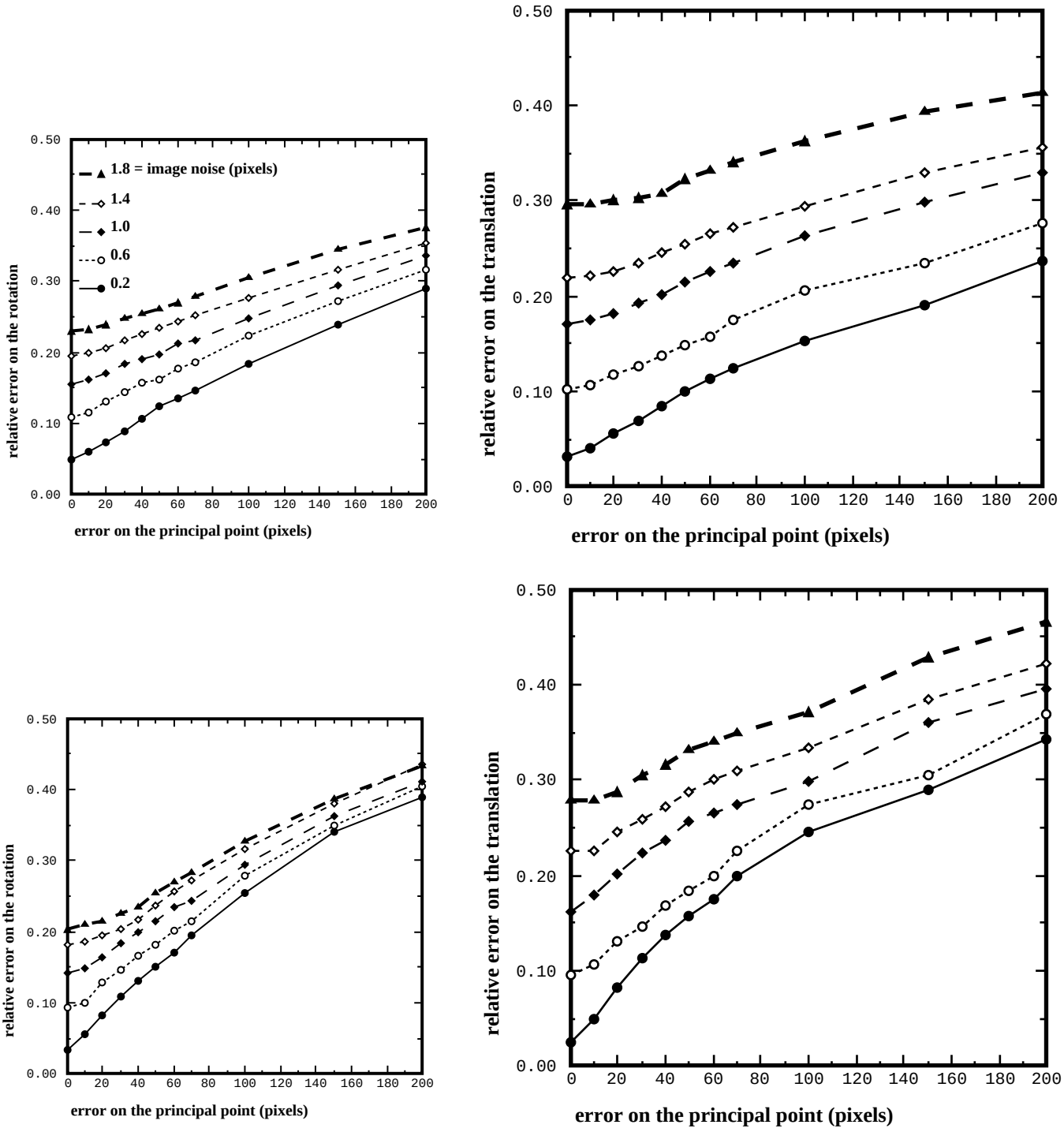

Figure 10: Sensitivity of motion computation to errors on the principal point. Top: **FACTOR**, Bottom: **MIN-DIST**, Left: rotation, Right: translation.

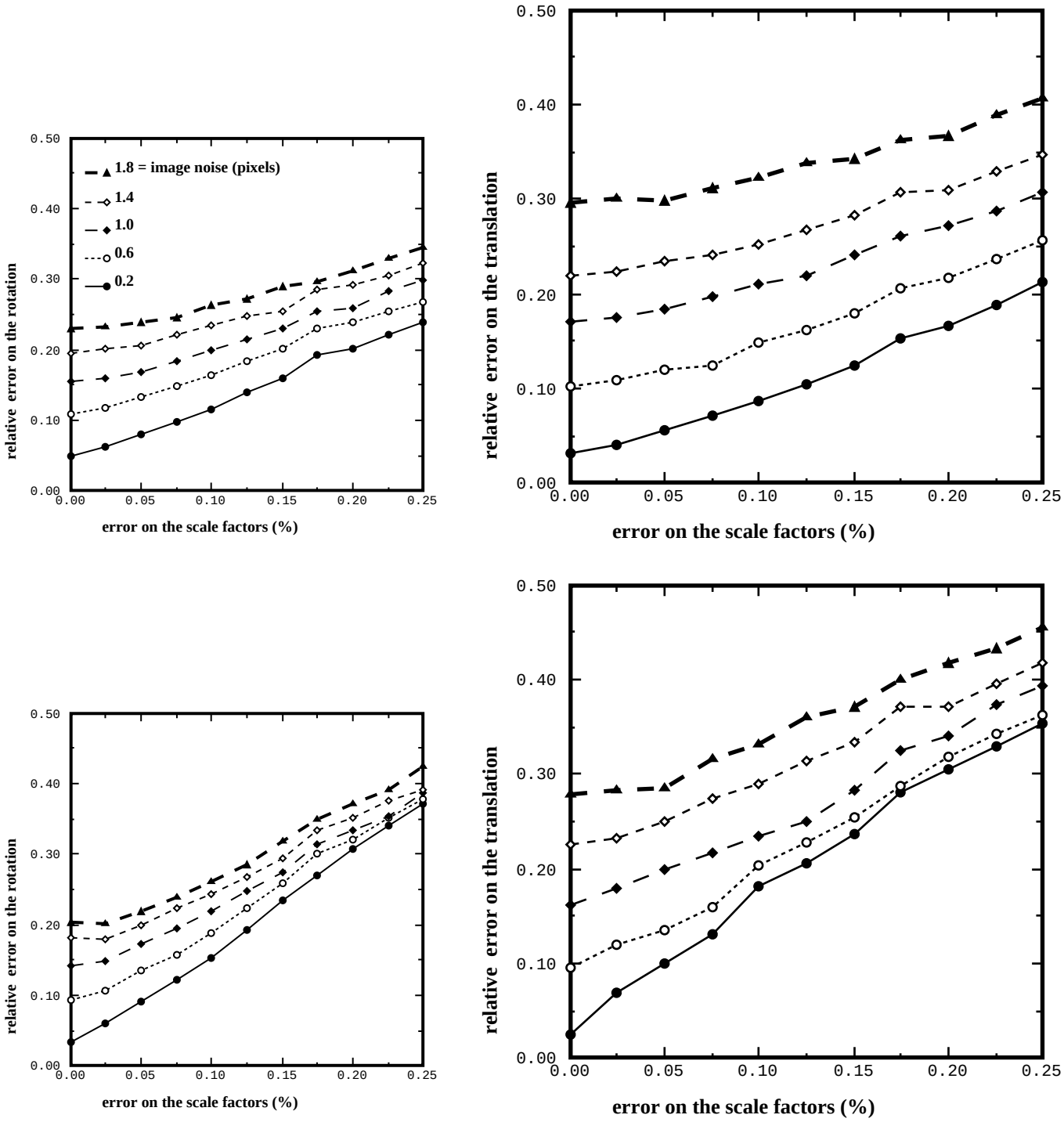

Figure 11: Sensitivity of motion computation to errors on the scale factors. Top: **FACTOR**, Bottom: **MIN-DIST**, Left: rotation, Right: translation.

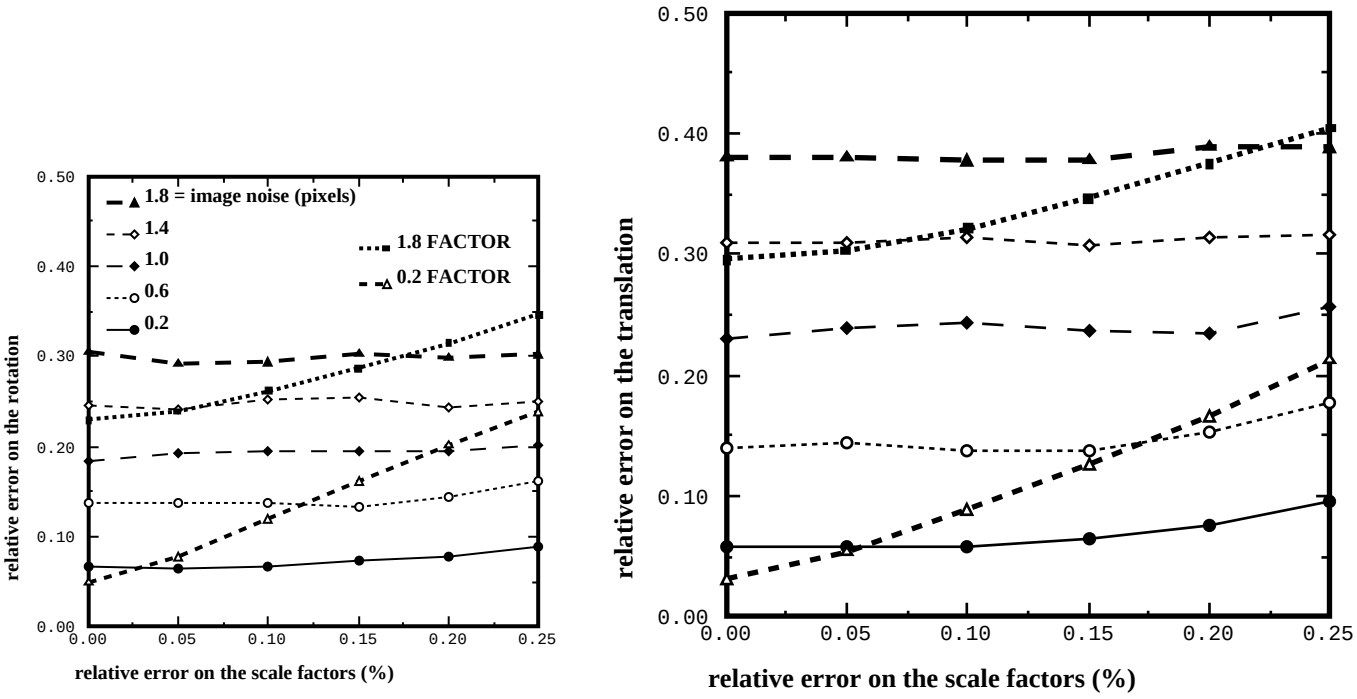

Figure 12: Sensitivity of motion computation to errors on the scale factors, in the case the scale factors are allowed to vary. Left: rotation, Right: translation.

Global computation of intrinsic parameters and motion

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

$$\min_{\substack{\alpha_u,\alpha_v,u_0,v_0 \\ \mathbf{r}_i,\frac{t_{ix}}{t_{iz}},\frac{t_{iy}}{t_{iz}}\ i=1\ldots N}} \sum_{i=1}^{N}\sum_{j} d^2(\mathbf{q}'_{ij}, \mathbf{A}^{-1T}\mathbf{T}_i\mathbf{R}_i\mathbf{A}^{-1}\mathbf{q}_{ij}) + d^2(\mathbf{q}_{ij}, \mathbf{A}^{-1T}\mathbf{R}_i^T\mathbf{T}_i\mathbf{A}^{-1}\mathbf{q}'_{ij}) \qquad (40)$$

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

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

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

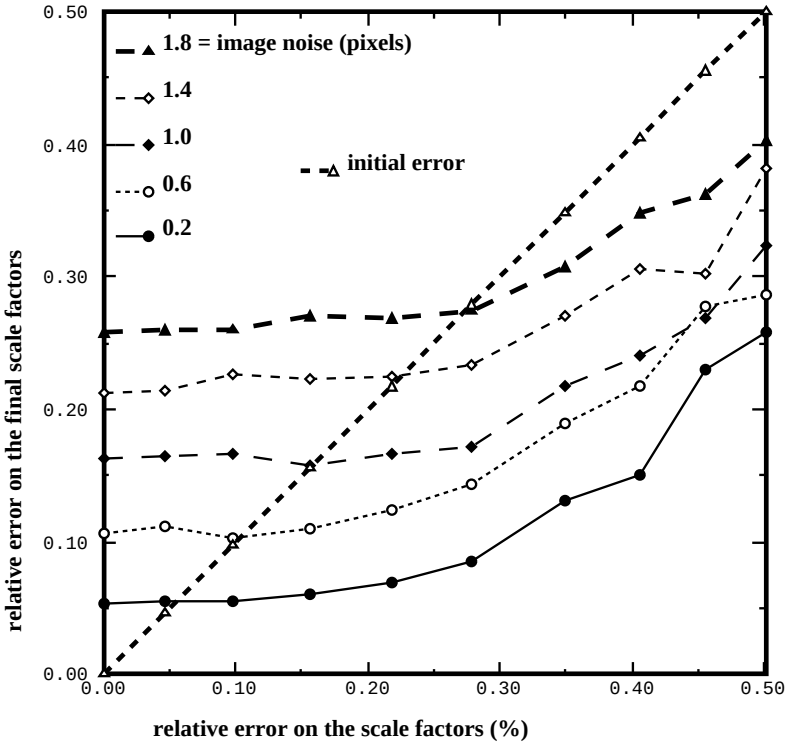

Figure 13: Final scale factors obtained when computing the motion.

Kruppa equations (denotated by **KRUPPA**). The denotations are followed by the number of displacements used. The abbreviation **MOUV** designates the global method, initialized with the starting point already used in the previous section (800, 800, 255, 255). The abbreviation **MOUV-KRUPPA** designates the global method, initialized with the values obtained by the Kruppa method. The image noise has the same meaning than previously, that is Gaussian noise added to pixel coordinates of point correspondences. In figure 14, 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. 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, the method based on the Kruppa equations 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 the global method based on the decomposition of the fundamental matrix, it is the *form of the parameterization* which ensures that all the constraints are satisfied. Then we have to compute explicitely 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 under the form $\mathbf{A}^{-1T}\mathbf{E}\mathbf{A}^{-1}$, 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 Kruppa equations, and these constraints cannot be enforced at the first stage of the computation. Moreover, using the technique mentionned later in Section 5.1 allows us to take into account trinocular constraints. Second, the criterion uses directly more information.

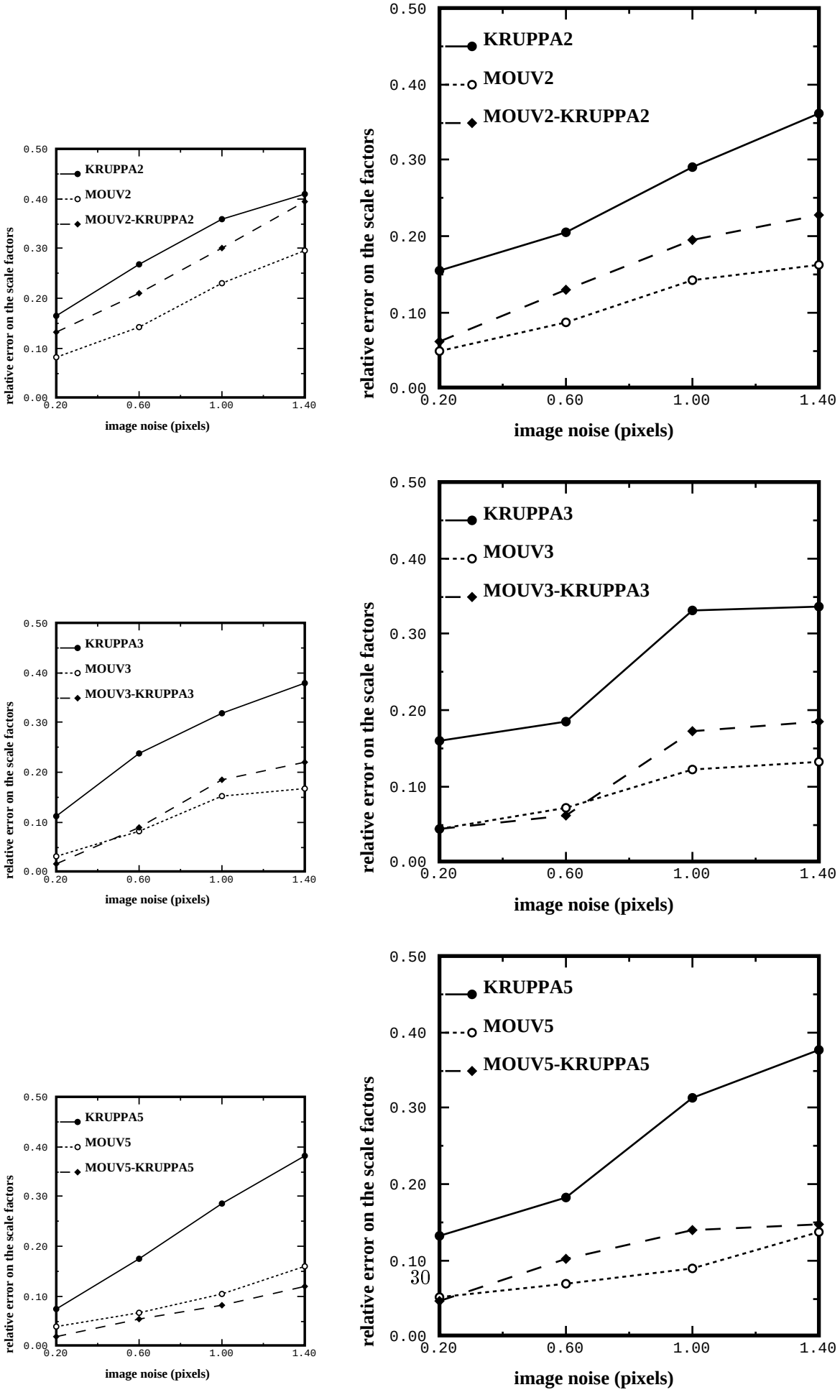

Figure 14: Comparison of the Kruppa based self-calibration method with the motion-based global

This explains that we obtain more precise results.

4.4 An evaluation of the results

In this section, we have studied the computation of the motion parameters in the context of self-calibration. One finding is that althougth 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. 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 perfectely adequate for this purpose, since some of them do not even need an initialization.

5 Computing the extrinsic parameters of a stereo rig and recovering structure from multiple viewpoints

In the usual calibration method, we work in the world coordinate system, using a 3D model of an object present in the environment. It is assumed that we know the 3D coordinates of some of reference points on the object in a coordinate system attached to the object. The extrinsic parameters define the displacement from the object's coordinate system (taken to be the world reference frame) to the camera coordinate system. In this part, we do not use any 3D model, so we do all the computations in the cameras coordinate system, and use as the world reference frame the one attached to the first camera. Thus, in our case, the extrinsic parameters define the displacement from the first to the second camera, computed in the first camera coordinate system. Since we do not have any metric information, we can compute this displacement only up to a scale factor.

Two different approaches are presented. The first one is straightforward in the case of a binocular stereo rig and more subtile in the case of a trinocular stereo rig, but it needs inter-camera point matching. The second enables us to obtain the inter-camera relative displacements using only monocular point matches. Two displacements of the stereo rig are, in general, sufficient to obtain a unique solution.

5.1 A direct approach: binocular and trinocular stereo rig

The most straightforward approach is to apply the techniques previously presented using point correspondences established between the different cameras of the stereo rig. How to obtain these correspondences automatically is not the subject of this paper. The advantage of this approach is that, since the relative displacement between the cameras is fixed, it is possible to accumulate point matches between pairs of images taken at different times. Using multiple displacements, it is possible to obtain a number of point matches far larger than the one that could be obtained from a single pair of images, which allows to obtain very precise results. This idea has been developped in [58]. Let us now explain how the perspective projection matrices are obtained, from available data which are now the relative displacements from cameras i to cameras j, obtained only up to a scale factor, and expressed in the coordinate system of camera i.

The binocular case The world coordinate system that is used all the way in this section is the coordinate system attached to the first camera, taken at the first position. It means, as mentionned in Section 2.1, that the perspective projection matrix attached to the first camera

can be written: $\mathbf{P}_1 = [\mathbf{A}_1 \ , \ 0]$, where $\mathbf{A}_1$ is the 3×3 intrinsic parameters matrix of the first camera. The extrinsic parameters of the stereo rig are the parameters of the displacement $\mathbf{D}_{12}$ between the first camera and the second camera, expressed in the coordinate system of the first camera. Thus the two projection matrices are:

$$\mathbf{P}_1 = [\mathbf{A}_1 \ , \ \mathbf{0}] \quad \mathbf{P}_2 = \mathbf{P}_1\mathbf{D}_{12} = [\mathbf{A}_2\mathbf{R}_{12}, \mathbf{A}_2\mathbf{T}_{12}] \tag{41}$$

where $\mathbf{A}_2$ is the intrinsic parameters matrix of the second camera, and $\mathbf{R}_{12}$ and $\mathbf{T}_{12}$ the rotational and translational components of the displacement $\mathbf{D}_{12}$. In practice, when working only from images, we only know the direction of the translation, and thus usually use the unit vector $\mathbf{t}_{12}$ instead of $\mathbf{T}_{12}$. If we replace $\mathbf{T}_{12}$ by $\mathbf{t}_{12}$ in (41), the last row of the projection matrix $\mathbf{P}_2$ is multiplied by an unknown scale factor. The 3D reconstruction obtained is nevertheless coherent, up to a scale factor. A single metric information concerning the motion, or the 3D length of a feature measured in an image, would be sufficient to determine this scale factor.

The trinocular case In the case of three cameras (designated by 1, 2 and 3), using again the first camera coordinate system as a reference, the three projection matrices can be written:

$$\mathbf{P}_1 = [\mathbf{A}_1 \ , \ \mathbf{0}] \quad \mathbf{P}_2 = \mathbf{P}_1\mathbf{D}_{12} \quad \mathbf{P}_3 = \mathbf{P}_2\mathbf{D}_{23} \tag{42}$$

Where $\mathbf{D}_{12}$ is in the coordinate system of the first camera, and $\mathbf{D}_{23}$ in the coordinate system of the second camera. These are quantities which can be computed from images, with the important detail that while computing camera motion, we are not able to determine the translations $\mathbf{T}_{12}$ and $\mathbf{T}_{23}$, but only the directions $\mathbf{t}_{12}$ and $\mathbf{t}_{23}$. While in the case of two views it is appropriate to use the formula (41) and to replace $\mathbf{T}_{12}$ by $\mathbf{t}_{12}$, since this results only in a global scene scale factor, if we want to perform the reconstruction using three views[11], we have to obtain *three coherent* projection matrices, and it is not appropriate to use formula (42) and to replace $\mathbf{T}_{12}$ and $\mathbf{T}_{23}$ by $\mathbf{t}_{12}$ and $\mathbf{t}_{23}$. This would yield an incorrect result, in which the epipolar constraint between the images 1 and 3 is not satisfied, since the direction of translation between $\mathbf{P}_1$ and $\mathbf{P}_3$ computed in this way would be generally incorect. The difficulty comes from the fact that if we know two displacements only up to a scale factor, it is only possible to determine the rotation:

$$\mathbf{R}_{13} = \mathbf{R}_{23}\mathbf{R}_{12} \tag{43}$$

but not the direction of the translation of $\mathbf{D}_{23}\mathbf{D}_{12}$, the only constraint being that it belongs to the plane: $\langle \mathbf{t}_{23}, \mathbf{R}_{23}\mathbf{t}_{12} \rangle$:

$$\mathbf{t}_{13} \cdot (\mathbf{t}_{23} \times \mathbf{R}_{23}\mathbf{t}_{12}) = 0 \tag{44}$$

The ratio $\lambda = \frac{\|\mathbf{t}_{12}\|}{\|\mathbf{t}_{23}\|}$ must be known, as well as the relative signs. Thus, from the knowledge of $\mathbf{R}_{12}$, $\mathbf{R}_{23}$, $\mathbf{t}_{12}$, $\mathbf{t}_{13}$ there is no way to build *three* coherent projection matrices. In order to determine the ratio, we have also be able to compute the displacement $\mathbf{D}_{13}$ in the first camera coordinate system, to determine the direction of translation $\mathbf{t}_{13}$. Since we want to reconstruct from the images 1, 2, and 3, there must be a portion of the scene visible in both image 1 and image 3, and thus, it is a reasonnable requirement. By expressing the proportionality constraint:

$$\mathbf{t}_{13} \times (\mathbf{R}_{23}\mathbf{u}_1 + \lambda\mathbf{u}_2) = 0 \tag{45}$$

where $\mathbf{u}_1 = \frac{\mathbf{t}_{12}}{\|\mathbf{t}_{12}\|}$ and $\mathbf{u}_2 = \frac{\mathbf{t}_{23}}{\|\mathbf{t}_{23}\|}$, we obtain:

$$\lambda = -\frac{(\mathbf{t}_{13} \times \mathbf{R}_{23}\mathbf{u}_1)_1}{(\mathbf{t}_{13} \times \mathbf{u}_2)_1} = -\frac{(\mathbf{t}_{13} \times \mathbf{R}_{23}\mathbf{u}_1)_2}{(\mathbf{t}_{13} \times \mathbf{u}_2)_2} = -\frac{(\mathbf{t}_{13} \times \mathbf{R}_{23}\mathbf{u}_1)_3}{(\mathbf{t}_{13} \times \mathbf{u}_2)_3} \tag{46}$$

Taking $\mathbf{t}_{12} = \mathbf{u}_1$, $\mathbf{t}_{23} = \lambda\mathbf{u}_2$, we then obtain, using (42) three perspective projection matrices that are all mutually coherent.

[11] It is well known that trinocular stereo algorithms are more efficient and yield more precise 3D reconstructions.

It can be noted that this approach can be also used in the case of structure from motion (eg: one unique moving camera) in order to register in a same coordinate frame an arbitrary number of positions. Further, we can use it to reduce the number of motion parameters to be estimated in the global approach to self-calibration of section 4 from three views. Instead of $5 \times 3 = 15$ motion parameters, we are left with $5 \times 2 + 1 = 11$ parameters which are the first two motions and the parameter λ defined in (46). This is a way to express the geometric constraints associated to a set of three views taken by the same camera[12],

5.2 An indirect, monocular approach

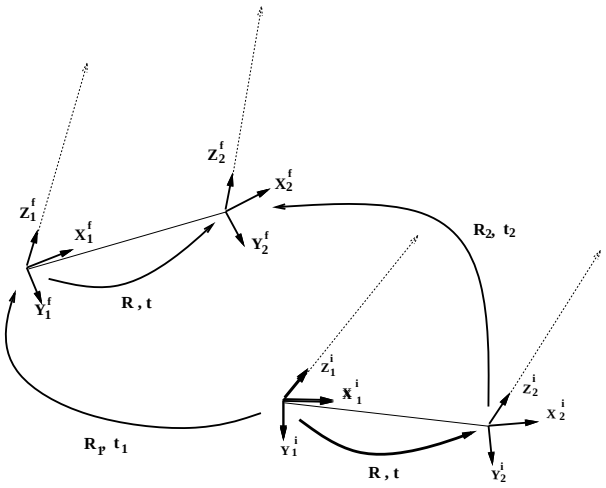

Figure 15: Displacement of a binocular stereo rig.

The principle This approach only requires to compute the displacement $\mathbf{D}_1$ of the first camera and the displacement $\mathbf{D}_2$ of the second camera, in the coordinate systems of the first and of the second cameras, respectively. The difficulty arises from the fact that $\mathbf{D}_1$ and $\mathbf{D}_2$ are then known in different coordinates systems, as shown in figure 15, in which the superscripts i and f refer to initial and final positions. To cope with this problem, the idea of our method is to use the commutativity of the following diagram:

$$\begin{array}{ccc} & C_1^f & \xrightarrow{\mathbf{D}} & C_2^f & \\ \mathbf{D}_1 & \uparrow & & \uparrow & \mathbf{D}_2 \\ & C_1^i & \xrightarrow{\mathbf{D}} & C_2^i & \end{array}$$

In this diagram, the relative displacement from the first to the second camera does not change between the initial and the final position of the stereo rig, since they are rigidly attached to each other. We can thus write the matrix equation:

$$\mathbf{D}\mathbf{D}_1 = \mathbf{D}_2\mathbf{D} \tag{47}$$

where $\mathbf{D}$ is the 4×4 unknown matrix of the displacement from the first camera to the second camera, $\mathbf{D}_1$, $\mathbf{D}_2$ are the 4×4 displacement matrices of the first and of the second cameras, in their initial respective coordinate systems. Equation (47) can be decomposed in the following two matrix equations:

$$\mathbf{R}\mathbf{R}_1 = \mathbf{R}_2\mathbf{R} \tag{48}$$

[12] This idea is considerably developped in [35].

$$(\mathbf{I} - \mathbf{R}_2)\mathbf{t} = \mu_2 \mathbf{t}_2 - \mu_1 \mathbf{R}\mathbf{t}_1 \tag{49}$$

where μ_1 and μ_2 are unknown scale factors associated to $\mathbf{D}_1$ and $\mathbf{D}_2$, respectively. The first equation has been much studied in the framework of hand-eye calibration [4] [44], [52], [3]. Thus the reader is refered to those references for a more detailed analysis of unicity and sensitivity to noise. We just indicate bellow how, if we perform two displacements of the stereo rig, we can solve the two resulting matrix equations (48) to compute $\mathbf{R}$, and point out to some constraints which arise from the displacement of a stereo rig. The solution of the two vector equations (49) to compute $\mathbf{t}$ up to a scale factor is less classical, since it involves working with translation vectors which are defined only up to a scale factor, as in the previous case of the trinocular stereo rig which it generalizes.

There is an important advantage of this method over the one which consists in computing directly the displacement from matches between the first and the second camera: since we work in each camera independently, we need only monocular matches which are more easy to obtain, since an arbitrary number of intermediate movements can be performed, and a token tracking procedure used. On the opposite, finding directly stereo matches can be difficult if the baseline of the stereo rig is large, since at this stage the stereo rig is not yet calibrated.

Recovering the rotation To solve equation (48), we use a quaternion representation of the rotations [40]

$$\begin{aligned} \mathbf{q}_\mathbf{R} &= (s, \mathbf{v}) \\ \mathbf{q}_{\mathbf{R}_1} &= (s_1, \mathbf{v}_1) \\ \mathbf{q}_{\mathbf{R}_2} &= (s_2, \mathbf{v}_2) \end{aligned}$$

Writing equation (48) with this representation yields:

$$\mathbf{q}_\mathbf{R} * \mathbf{q}_{\mathbf{R}_1} - \mathbf{q}_{\mathbf{R}_2} * \mathbf{q}_\mathbf{R} = 0 \tag{50}$$

where $*$ indicates the quaternion product. This gives the two equations:

$$\begin{aligned} s(s_1 - s_2) &= \mathbf{v}.(\mathbf{v}_1 - \mathbf{v}_2) \\ s(\mathbf{v}_1 - \mathbf{v}_2) + (s_1 - s_2)\mathbf{v} + \mathbf{v} \times (\mathbf{v}_1 + \mathbf{v}_2) &= 0 \end{aligned} \tag{51}$$

Let us write: $\mathbf{v} = \alpha \mathbf{v}_1 + \beta \mathbf{v}_2 + \gamma(\mathbf{v}_1 \times \mathbf{v}_2)$. After some algebra using the properties of the quaternions, we obtain:

$$\alpha = \beta \quad , \quad s_1 = s_2 \tag{52}$$

and then:

$$s = \gamma(\mathbf{v}_1^2 + \mathbf{v}_1.\mathbf{v}_2)$$

$$\alpha^2 \|\mathbf{v}_1 + \mathbf{v}_2\|^2 + \gamma^2(\|\mathbf{v}_1 \times \mathbf{v}_2\|^2 + \frac{\|\mathbf{v}_1 + \mathbf{v}_2\|^4}{4}) = 1 \tag{53}$$

Since there are two unknowns α and γ, t his last equation determines a one-parameter family of rotations, which are parameterized by an ellipse lying in the plane $(\mathbf{v}_1 + \mathbf{v}_2, \mathbf{v}_1 \times \mathbf{v}_2)$. It has been shown ([4], an alternative proof is in [31]) that a necessary and sufficient condition for a unique solution is that two displacements with non-parallel rotation axis be performed. Let us point out to a binocular constraint which is present in our case:

$$\begin{aligned} \mathbf{R}\mathbf{R}_1 &= \mathbf{R}_2\mathbf{R} \\ \mathbf{R}\mathbf{R}'_1 &= \mathbf{R}'_2\mathbf{R} \end{aligned}$$

From these relations, we see that we have also:

$$\mathbf{R}(\mathbf{R}_1\mathbf{R}'_1) = (\mathbf{R}_2\mathbf{R}'_2)\mathbf{R}$$

Using the constraint (52) on the last equation yields: $s_1 s'_1 - \mathbf{u}_1.\mathbf{u}'_1 = s_2 s'_2 - \mathbf{u}_2.\mathbf{u}'_2$. Since $s_1 = s_2$ and $s'_1 = s'_2$, we finally obtain the equation:

$$\mathbf{u}_1.\mathbf{u}'_1 = \mathbf{u}_2.\mathbf{u}'_2 \tag{54}$$

For practical purposes, instead of using (53), we can obtain, as in [4], a linear solution. We can notice that (50) has the same form as (37). Thus there exists a 4×4 matrix $\mathbf{G}$, such that:

$$\mathbf{q}_\mathbf{R} \times \mathbf{q}_{\mathbf{R}_1} - \mathbf{q}_{\mathbf{R}_2} \times \mathbf{q}_\mathbf{R} = \mathbf{G}\mathbf{q}_\mathbf{R} \tag{55}$$

$\mathbf{G}$ is given by a formula similar to (37). A closed form solution can be obtained with two equations (55) obtained by two displacements of the stereo rig. If we use more displacements, we can improve the results by using a linear least-squares procedure.

Recovering the direction of the translation We suppose that we have computed $\mathbf{R}$, as previously described. A geometrical analysis shows that the matrix $\mathbf{I} - \mathbf{R}_2$ maps all vectors in the plane perpendicular to the axis $\mathbf{u}_2$ of the rotation $\mathbf{R}_2$. Thus, starting from relation (49), we can write:

$$\mathbf{u}_2.(\mu_2\mathbf{t}_2 - \mu_1\mathbf{R}\mathbf{t}_1) = 0$$

This allows us to determine the ratio $a = \mu_1/\mu_2$. It is then possible to recover the direction $\mathbf{t}_-$ of the component of $\mathbf{t}$ orthogonal to $\mathbf{u}_2$, yielding the constraint:

$$\mathbf{t} \in \langle \mathbf{t}_-, \mathbf{u}_2 \rangle \tag{56}$$

If a second movement, *for which the axis* $\mathbf{u}'_2$ *of the rotation is different*, is performed, we can compute similarly a direction $\mathbf{t}'_-$. Combining the two constraints (56), and the same with primes, we obtain $\mathbf{t}$ up to a scale factor by:

$$\mathbf{t} = \lambda(\mathbf{t}_- \times \mathbf{u}_2) \times (\mathbf{t}'_- \times \mathbf{u}'_2) \tag{57}$$

Note that if we perform more than two displacements, the direction of $\mathbf{t}$ can be easily recovered by using a linear least-squares procedure based on equation (57). This completes the computation of the relative position of the two cameras, up to a scale factor.

6 Experimental results

6.1 An example of calibration of a binocular stereo rig

Self-calibration of a camera We first show the results of the monocular self-calibration, using three images taken by the left camera at different positions of the stereo rig. Results are quite similar for the second camera. 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 in the left camera.

We use between 20 and 30 corners, which are extracted with a sub-pixel accuracy, semi-automatically, by the program of T. Blaszka and R. Deriche [7]. Correspondence is, in this experiment, performed manually, and followed by an automatic elimination of false matches. 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 16 shows the points of interest matched between image 1 and image 2. The standard calibration is performed on each image, using the algorithm of Robert [41], which is a much improved version of the linear method of Faugeras and Toscani [15]. 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. 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 (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, 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 value remains under one pixel. Some epipolar lines obtained with points that are seen in the three images are shown figure 17 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. It can be noted that no initial guess is required at all for the general method. 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. 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.

Figure 16: A pair of images with the detected corners superimposed.

Extrinsic parameters computation Once the self-calibration of each camera has been achieved, we have performed two other displacements of the stereo rig. We have not used

	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

Table 6: Results of the fundamental matrix estimation in the left camera.

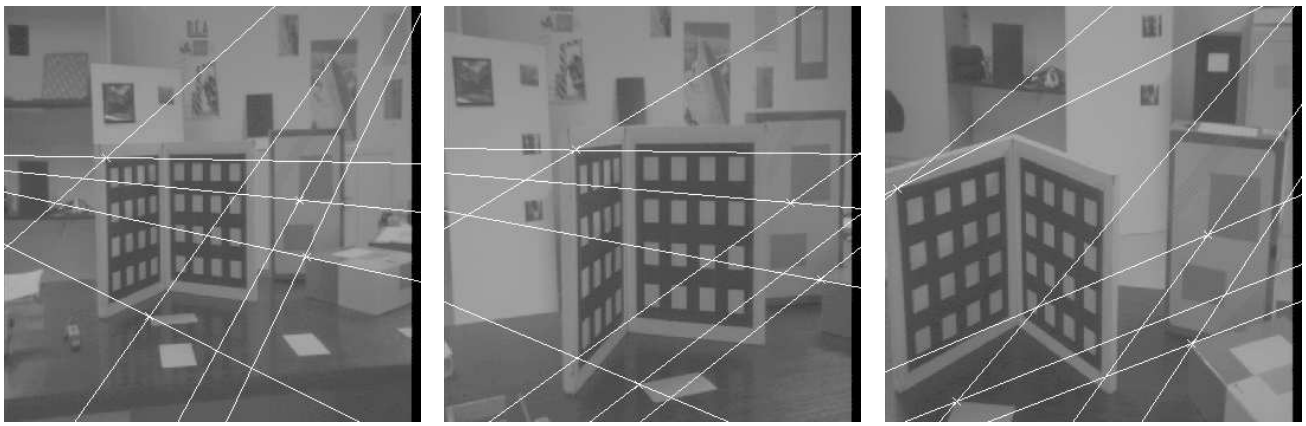

Figure 17: A triplet of images with some estimated epipolar lines superimposed.

method	α_u	α_v	u_0	v_0	$\theta - \frac{\pi}{2}$
grid, image 1	657.071	1003.55	244.227	256.617	-2.05e-06
grid, image 2	664.975	1015.2	232.61	257.701	-7.47e-07
grid, image 3	639.749	980.185	252.174	249.585	-2.60e-06
Kruppa polynomial	639.405	982.903	258.980	341.013	-6.11e-03
Kruppa iterative	640.12	936.08	206.17	284.95	-0.07
Kruppa iterative (center)	681.28	985.69	255	255	0

Table 7: Results of the intrinsic parameters estimation in the left camera.

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 8: Results of the camera motion estimation in the left camera (first sequence).

1-2	r_1=|-0.00012, 0.3130, 0.00773|^T	t_1 = [0.1237, -0.0209, 0.9920]^T	α_1 = .3131
	r_2 =[.00554, .31196, -.01219]^T	t_2 = [0.2953, 0.0160 0.9552]^T	α_2 = .3122
2-3	r'_1 = [-0.0334, -0.1098, -0.143]^T	t'_1 = [-0.124, 0.5974, 0.7922]^T	α'_1 = .1833
	r'_2 = [-0.0540, -0.117, -0.133]^T	t'_2 = [-0.01089, 0.02439,0.9996]^T	α'_2 = .1853
1-3	r''_1 = [-0.00224, 0.2054, -0.1307]^T	t''_1 = [-0.1882, 0.9809, 0.0476]^T	α''_1 = .2435
	r''_2 = [-0.06175, 0.198, -0.1385]^T	t''_2 = [0.3423, 0.1487,0.9277]^T	α''_2 = .2494

Table 9: Results of the camera motion estimation in the left and right camera. (second sequence).

the three previous displacements because they yield computations that are less stable for the method we want to illustrate now: the computation of the relative displacement between the two cameras of the rig using only monocular matches, the problem being the little difference of motion between the two cameras of the rig. We have performed small displacements which maximize this difference. The six images are shown in figure 18 Since only a small part of the calibration grid is seen, we cannot directly check the results of the determination of camera motion shown in the table 9. However, we verify the consistency of these results thanks to two families of constraints: the one arising from the fact that the two cameras of the rig are rigidly attached, and the one arising from the fact that the third displacement is a composition of the first two displacements, since only three *images* are used. The binocular constraints are that the angles of rotations of the two cameras are equal for a given displacement of the rig (52), which can be checked in the last column of table 9 and the relation (54), whose residual values are here $-.00943$, .01950, and $-.05394$. The monocular constraints are obtained from the fact that the composition of the two first motions gives the third one. We obtain for the rotations, using (43):

$$r_1^{''} = [-.01125, .2027, -.1389]^T \quad r_2^{''} = [-.02648, .1936, -.1518]^T$$

which is close to the values actually computed shown in table 9: the relative error on the angles is 1% and 0.8%, and the angle between the axes is 8.7^o and 2.9^o. The triple product (44) involves also the direction of translations. The value in the left camera is $-.0192$, and in the second camera $-.002$. Thus we have checked that all the constraints are well satisfied.

We have then computed the relative displacement between the two cameras of the rig, using different methods:

- The classical calibration method. The reference position is taken in such a way that the grids cover a large part of each image. It can be noted that when using other positions (the first two positions used for self-calibration, where the grids can be seen entirely), the results vary significantly.
- The direct method using stereo matches. It yields very stable results. Adding correspondences through images improves rotation accuracy.
- The indirect method, using the three pairs of motions, gives results comparable to those obtained with images where calibration grids do not "fill the image frame".

The results are in table 10, which shows the rotation vector and the normalized translation vector, as well as the relative error on the rotation angle and the angular error and the rotation axis and translation direction. Thus, good results can be obtained if stereo correspondences are available, and reasonable results are obtained by the monocular method. Precision can be easily improved by using more images than the minimal number used here.

6.2 Varying the focal length

We have applied the method to a camera with a variable focal length. The results are shown in table 11. It allows us to notice that the best results are obtained for short focal lengths, which

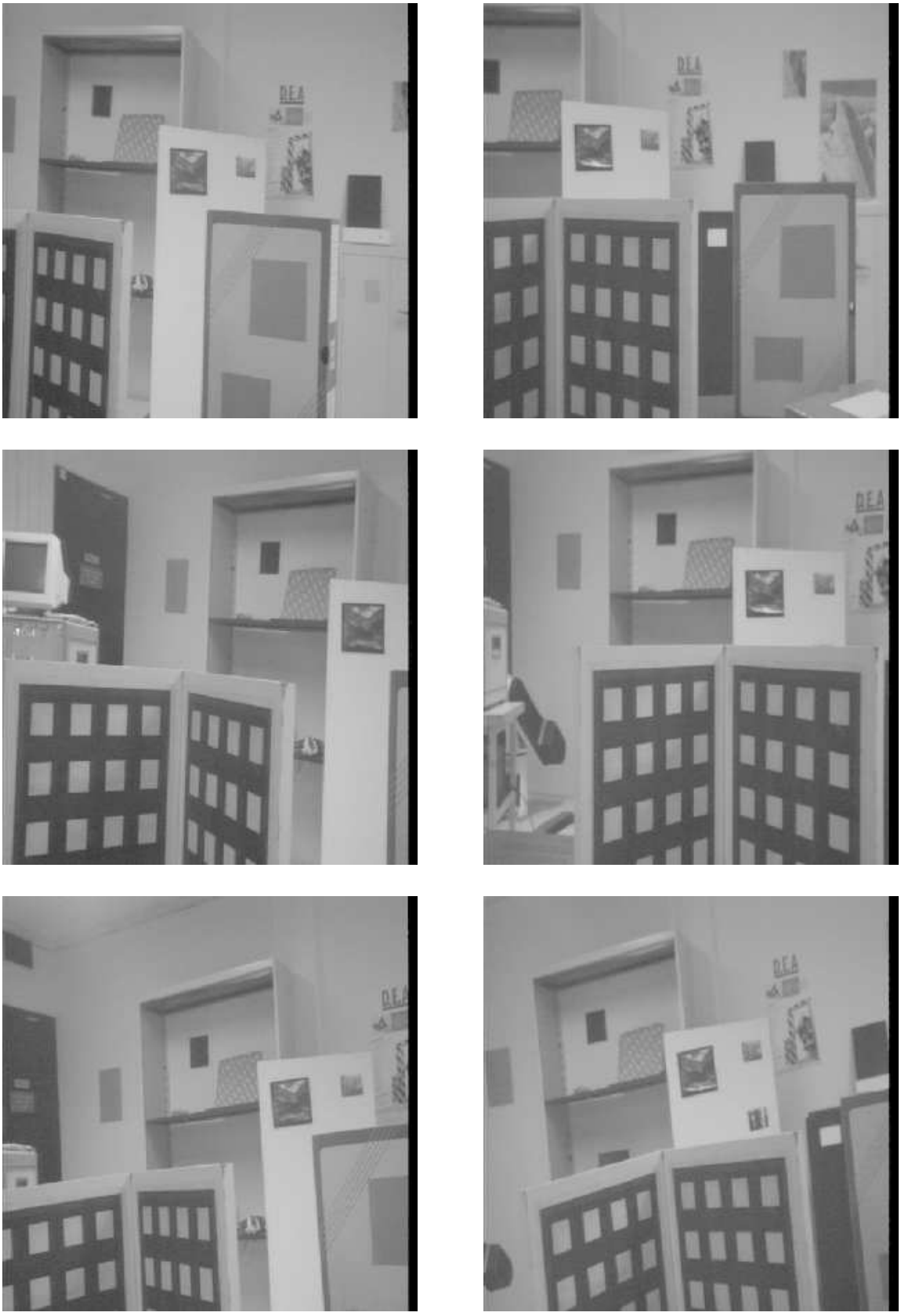

Figure 18: Three pairs of stereo images (arranged for cross-viewing).

method	r_x	r_y	r_z	t_x	t_y	t_z	$\frac{\Delta\alpha}{\alpha}$	θ_r	θ_t
GRID (ref.)	-0.04097	0.1842	0.05561	0.9992	0.03770	-0.00889			
GRID (1)	-0.04015	0.2285	0.05573	0.9970	0.04541	0.0613	0.21	3.8	4.0
GRID (2)	-0.03595	0.2042	0.05611	0.9976	0.04234	0.05303	0.09	2.7	3.5
STEREO (1)	-0.03383	0.2205	0.05298	0.9992	0.03879	-0.00402	0.16	4.7	0.28
STEREO (2)	-0.03014	0.2025	0.05411	0.9992	0.03902	-0.00157	0.07	4.2	0.42
STEREO (1 + 2)	-0.04307	0.1895	0.05411	0.9991	0.03804	-0.01655	0.025	0.9	0.44
MONO	-.04915	.2383	.05322	.9987	-.002223	-.04902	0.26	4.1	3.2

Table 10: Results of the estimation of the relative displacement between the two cameras.

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

Table 11: Parameters obtained with a zoom camera

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.

6.3 Reconstructions from a triplet of uncalibrated images taken by a camera

We now show examples of reconstruction using structure from from motion with three uncalibrated views. The approach is to use the global minimization approach presented in Section 4.3, with the variant presented Section 5.1 to account for trinocular constraints.

A qualitative experiment The first set of images is used to illustrate the feasability of the method in a fairly standard indoor environment, such as it appears in the three views of figure 19. First, self-calibration is performed using the same method as in the previous examples. Then edge detection is performed, and the edge chains are approximated by B-splines, which are provided as input to the trinocular stereovision algorithm of Luc Robert [42, 41]. The matching phase of this algorithm uses only the epipolar geometry obtained from the fundamental matrices $\mathbf{F}_{12}$, $\mathbf{F}_{13}$, and $\mathbf{F}_{23}$, which are computed from the point correspondences. The 3D reconstruction phase requires in addition three projection matrices which relate the three image coordinate systems to a common world coordinate system. They are obtained by taking as the world coordinate system, the first camera coordinate system, and by finding the two displacements $\mathbf{D}_{12}$, $\mathbf{D}_{23}$, as well as the ratio of the norms of $\mathbf{t}_{12}$ and $\mathbf{t}_{23}$ (for which the computation of $\mathbf{D}_{13}$ is needed), as explained in section 5.1. Results of the reconstruction are shown in figure 20 as a stereogram which shows that planar structures and angles are quite well captured. The figure 21 shows two rotated views of the reconstructed scene, one from the side, the other from the top. It can be seen on these views that the estimated distances are also metrically plausible.

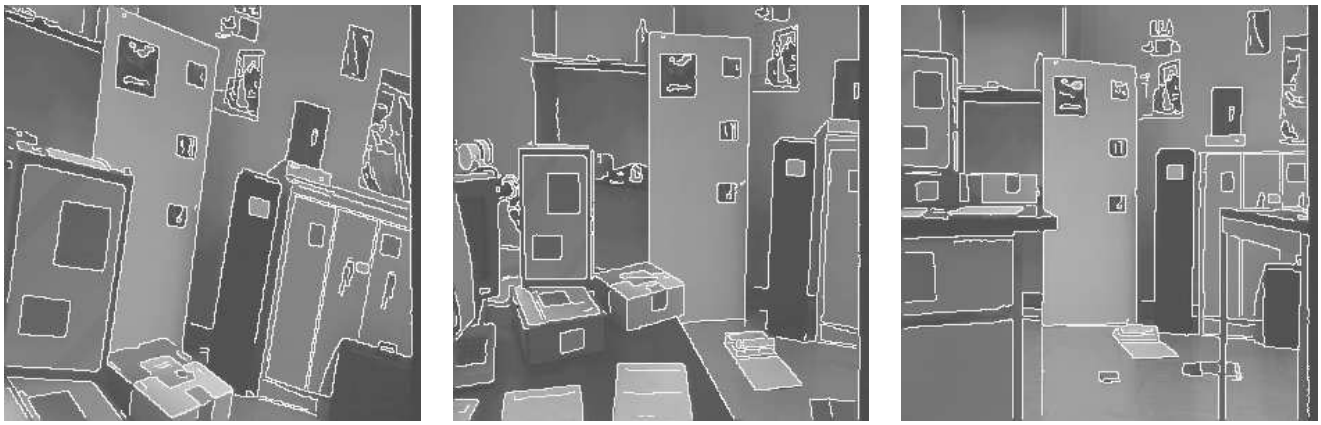

Figure 19: The triplet of images of the indoor scene, with edge chains superimposed.

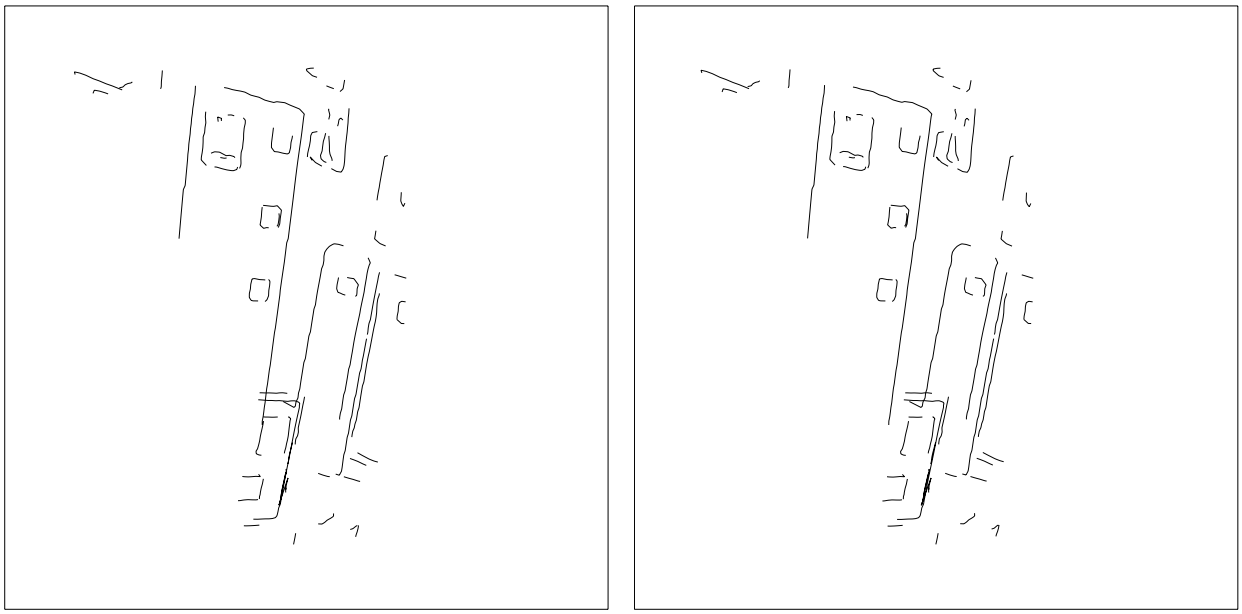

Figure 20: Reconstruction of the indoor scene from the uncalibrated triplet (stereogram for cross-viewing).

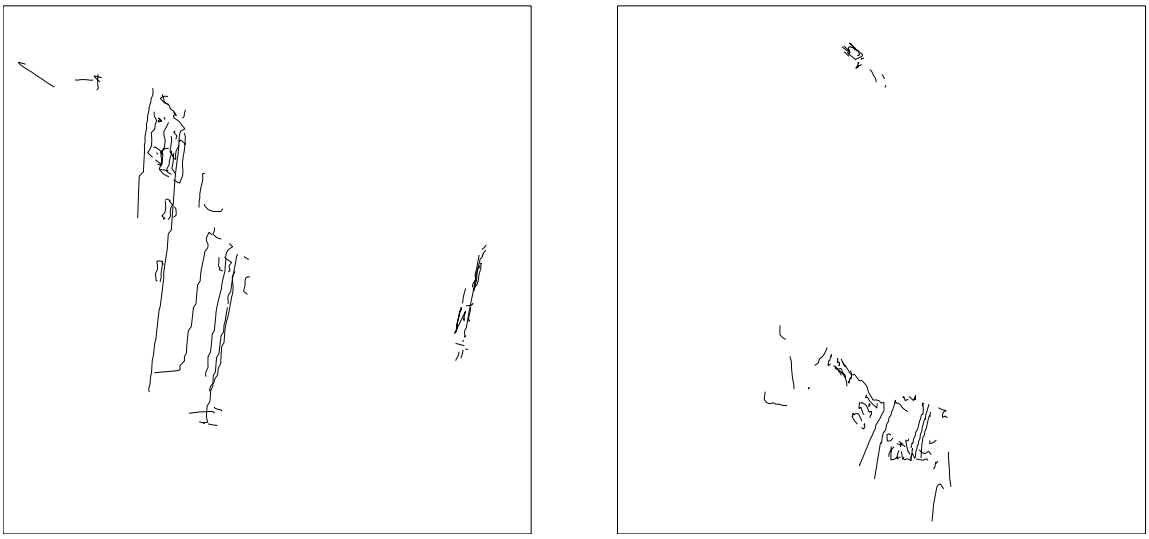

Figure 21: Two rotated views of the indoor scene.

A quantitative experiment We have further 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. In contrast with the previous images, 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 figure 22. The points of interest are the light dots and have been located and matched manually[13]. 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 [31, 33]. However, the epipolar geometry found from the three projection matrices obtained by self-calibration is fairly coherent, as illustrated in figure 23, which shows a zoom with epipolar lines of one the point of interest. 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 R. Vaillant and R. Deriche [8]. 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 Z. Zhang. After applying this similarity to the initial reconstruction, the final average error in 3D space with this sequence is 2 millimeters[14]. A sample of coordinates of reconstructed points are shown in table 12, units being in millimeters. It can be shown that the precision is about 1 part in 50.

7 Conclusion

We have presented a general framework to perform the self-calibration of systems of one, or several cameras. The basic idea is that the only information which is needed to perform calibration are point correspondences. This is contrast with all standard calibration methods. As a side effect of the calibration procedure, we can also estimate the relative displacements between

[13] A snake-based ellipse localization program due to B. Bascle, has also been tried.

[14] This is typical, more precise results have been sometimes achieved.

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	0.035	-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

Table 12: Comparison of the 3D reconstruction from self-calibration with reference points.

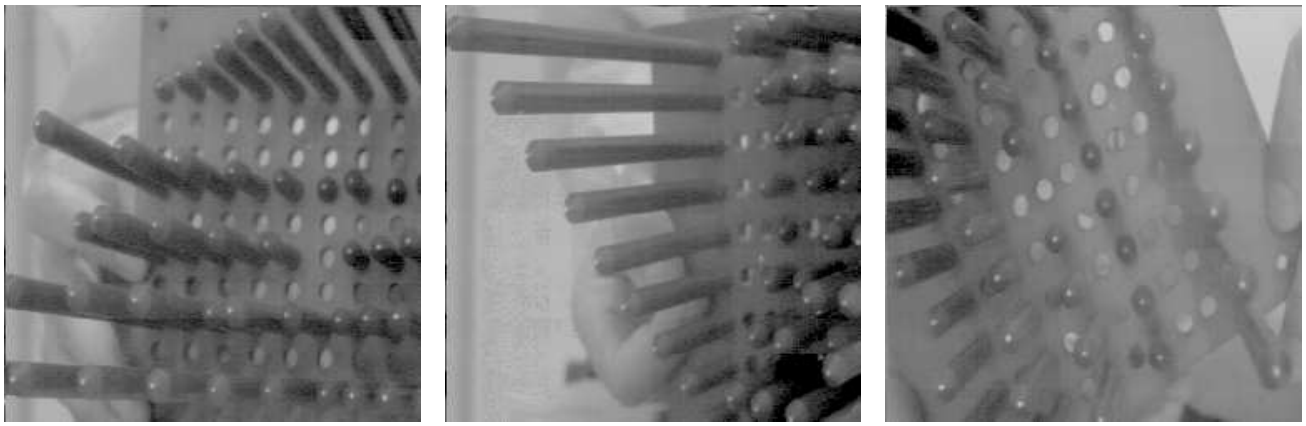

Figure 22: The triplet of images of the photogrammetric object.

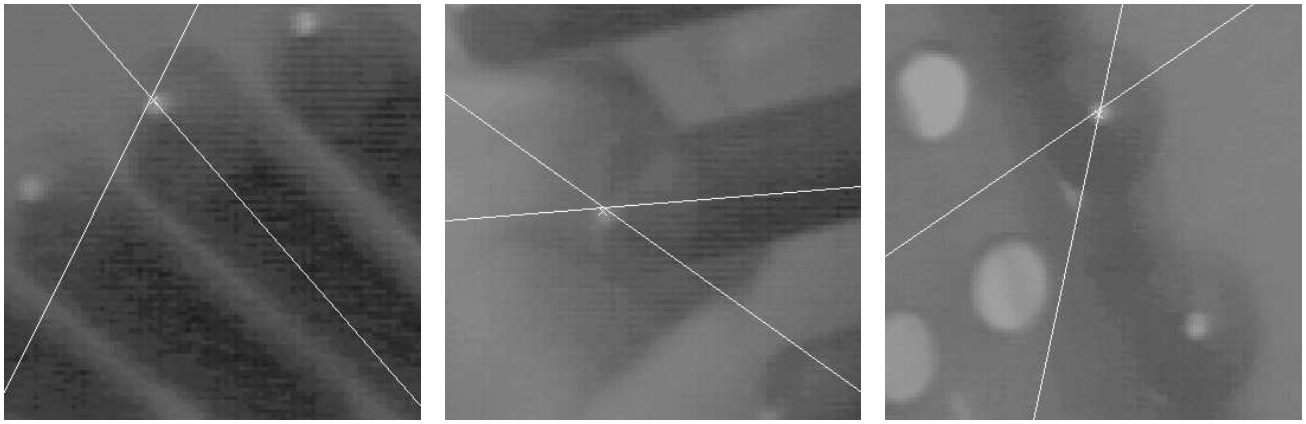

Figure 23: Zoom on the photogrammetric triplet, showing corresponding epipolar lines.

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. Thus a number of researchers have recently investigated self-calibration techniques. However, 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 [6], or restriction on the camera motions [9, 2, 27, 54]. 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, the most general work is that of Viéville [54] where the only additional assumption is the fact that the motion is a fixed-axis rotation, something well-suited to robotics heads. More precise and robust results are then obtained.

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 than 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 localisation and reliable correspondences are necessary.

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 (section 5). 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.

Acknowledgements

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 connecting the cameras.

A A few proofs of equivalence

A.1 Trivedi equations and 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 (14). Then follows immediately $\det(\mathbf{E}\mathbf{E}^T) = 0$, and thus the first condition $\det(\mathbf{E}) = 0$ is satisfied. Adding T_{12}, T_{13} and T_{23}, yields :

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

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:

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

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 $\mathbf{\Theta}$ such as:

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

This matrix equality can be expanded as:

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

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 exactely the form (14).

A.2 Huang-Faugeras constraints and 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:

$$\mathbf{F} = \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \mu & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{58}$$

where $\lambda \neq 0$ and $\mu \neq 0$ since we know that a fundamental matrix must be of rank two. Using (58) we obtain easily the epipoles $\mathbf{e} = \mathbf{e}' = (0,0,1)^T$ and the homography $h \ : \tau \mapsto -\frac{\lambda}{\mu}\tau$, and then, after some algebra, the Kruppa equations:

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

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):

$$\begin{aligned} \delta_3 &= \langle \mathbf{l}_1, \mathbf{l}_2 \rangle \\ \delta_{13} &= -\|\mathbf{l}_2\|^2 \\ \delta_{23} &= -\|\mathbf{l}_1\|^2 \end{aligned} \tag{59}$$

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

$$\lambda\delta_{23}E_1 - \mu\delta_{13}E_2 = \delta_3 E_3 \tag{60}$$

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:

$$\begin{aligned} f(\mathbf{E}) &= -\frac{1}{2}((\lambda^2\delta_{23}\delta'_{23} - \mu^2\delta_{13}\delta'_{13})^2 + 2\lambda\mu(\lambda\delta_3\delta'_{23} + \mu\delta_{13}\delta'_3)(\lambda\delta_{23}\delta'_3 + \mu\delta_3\delta'_{13})) \\ &= -\frac{1}{2}(E_3^2 + 2\lambda\mu E_1 E_2) \end{aligned}$$

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 (60):

$$(\lambda\delta_{23}E_1 - \mu\delta_{13}E_2)^2 + 2\lambda\mu E_1 E_2 \delta_3^2 = 0 \tag{61}$$

Thus:

$$\lambda^2\delta_{23}^2E_1^2 + \mu^2\delta_{13}^2E_2^2 = 2\lambda\mu E_1E_2(\delta_{13}\delta_{23} - \delta_3^2) \tag{62}$$

According to the definitions (59) 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 (62) 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 equation (61) 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:

$$\left\{ \begin{array}{l} \delta'_3 = 0 \\ \lambda^2\delta_{23}\delta'_{23} - \mu^2\delta_{13}\delta'_{13} = 0 \end{array} \right. \quad \text{ou} \quad \left\{ \begin{array}{l} \delta'_3 \neq 0 \\ \delta_{13} = \delta_{23} = 0 \end{array} \right.$$

and to:

$$f(\mathbf{E}) = 2\lambda^2\mu^2\delta_{13}\delta_{23}\delta_3'^2 + (\lambda^2\delta_{23}\delta'_{23} - \mu^2\delta_{13}\delta'_{13})^2 = 0$$

B Independence of Kruppa equations from three images

The two first displacements are:

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

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

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

$$\mathbf{R}_3 = \mathbf{R}_1\mathbf{R}_2 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & -1 \\ -1 & 0 & 0 \end{bmatrix} \qquad \mathbf{t}_3 = \mathbf{R}_1\mathbf{t}_2 + \mathbf{t}_1 = \begin{bmatrix} 3 \\ 3 \\ 1 \end{bmatrix}$$

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:

$$\begin{aligned}
E_1 &= 3\,\delta_1 - 2 + 6\,\delta_3\,\delta_{13} + 9\,\delta_3 + 4\,\delta_1\,\delta_3 - 7\,\delta_2\,\delta_{13} + 2\,\delta_2 + 12\,\delta_1\,\delta_2 + 3\,\delta_1\,\delta_{13} + 2\,{\delta_{13}}^2 \\
E'_1 &= 3\,\delta_{23} - 3\,\delta_{13}\,\delta_{23} + 8\,\delta_1\,\delta_{23} + 1 + \delta_1 \\
&\quad - \delta_2\,\delta_{13} - 4\,\delta_2 - 4\,\delta_1\,\delta_2 - 4\,\delta_3\,\delta_{13} - \delta_3 + 4\,\delta_1\,\delta_3 - {\delta_{13}}^2 + \delta_1\,\delta_{13} \\
E_2 &= 2\,\delta_3\,\delta_{23} + 16\,\delta_3 - 8\,\delta_2\,\delta_3 + 4\,\delta_2\,\delta_{23} \\
&\quad + 16\,\delta_2 - 16\,{\delta_2}^2 + 4\,\delta_1\,\delta_{13} + 16\,\delta_1 - 16\,{\delta_1}^2 + 2\,\delta_3\,\delta_{13} - 8\,\delta_1\,\delta_3 \\
E'_2 &= {\delta_{23}}^2 + 4\,\delta_{23} - 4\,\delta_2\,\delta_{23} - {\delta_{13}}^2 - 4\,\delta_{13} + 4\,\delta_1\,\delta_{13} \\
E_3 &= 6\,\delta_{23} + 6\,\delta_3 + 18\,\delta_{23}\,\delta_3 + 12\,\delta_3\,\delta_{13} + 36\,{\delta_3}^2 + 18\,\delta_{23}\,\delta_2 + 36\,\delta_2\,\delta_{13} + 36\,\delta_2\,\delta_3 \\
&\quad - 6\,\delta_{23}\,\delta_1 - 12\,\delta_1\,\delta_{13} + 18\,\delta_2 - 36\,\delta_1\,\delta_2 - 6\,{\delta_{13}}^2 - 18\,\delta_1 + 36\,{\delta_1}^2 \\
E'_3 &= 9\,{\delta_{23}}^2 + 9\,\delta_{23}\,\delta_{13} + 18\,\delta_{23}\,\delta_3 + \delta_{23} - 9\,\delta_{13} + 2\,\delta_3 + 6\,\delta_{23}\,\delta_2 + 6\,\delta_2\,\delta_{13} + 12\,\delta_2\,\delta_3 \\
&\quad - {\delta_{13}}^2 - 4\,\delta_1\,\delta_{13} - 9 + 12\,\delta_1 + 12\,{\delta_1}^2
\end{aligned}$$

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:

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

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.

References

[1] N. Ayache. *Vision stereoscopique et perception multisensorielle.* InterEditions, 1989.

[2] A. Basu. Active calibration: alternative strategy and analysis. In *Proc. of the conf. on Computer Vision and Pattern Recognition*, pages 495–500, New-York, 1993.

[3] H.H. Chen. A screw motion approach to uniqueness analysis of head-eye geometry. In *Proc. of the conf. on Computer Vision and Pattern Recognition*, pages 145–151, 1991.

[4] J.C.K. Chou and M. Kamel. Quaternions approach to solve the kinematic equation of rotation, $A_aA_x = A_xA_a$, of a sensor-mounted robotic manipulator. In *Proc. International Conference on Robotics and Automation*, pages 656–662, 1988.

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

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

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

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

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

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

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

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

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

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

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

[16] Olivier D. Faugeras, Francis Lustman, and Giorgio Toscani. Motion and Structure from point and line matches. In *Proc. International Conference on Computer Vision*, pages 25–34, June 1987.

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

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

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

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

[21] A.M. Jazwinsky. *Stochastic processes and filtering theory.* Academic Press, London, 1970.

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

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

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

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

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

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

[28] Yuncai Liu and Thomas S. Huang. A Linear Algorithm for Motion Estimation Using Straight Line Correspondences. *Computer Vision, Graphics, and Image Processing*, 44:35–57, 1988.

[29] Yuncai Liu and Thomas S. Huang. Estimation of Rigid Body Motion Using Straight Line Correspondences. *Computer Vision, Graphics, and Image Processing*, 43:37–52, 1988.

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

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

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

[33] Q.-T. Luong and O.D. Faugeras. The fundamental matrix: theory, algorithms, and stability analysis. Submitted, 1993.

[34] Q.-T. Luong and O.D. Faugeras. Fundamental matrix and self-calibration in computer vision. In preparation, 1994.

[35] Q.-T. Luong and T. Viéville. Canonic representations for the geometries of multiple projective views. Technical Report UCB/CSD-93-772, University of California at Berkeley, Sept 1993.

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

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

[38] P.S. Maybeck. *Stochastic models, estimation and control.* Academic Press, London, 1979.

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

[40] E. Pervin and J.A. Webb. Quaternions in computer vision and robotics. In *Proc. International Conference on Computer Vision and Pattern Recognition*, pages 382–383, 1983.

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

[42] L. Robert and O.D. Faugeras. Curve-Based Stereo: Figural Continuity And Curvature. In *Proc. International Conference on Computer Vision and Pattern Recognition*, pages 57–62, Maui, Hawai, June 1991. IEEE.

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

[44] Y.S. Shiu and S. Ahmad. Calibration of wrist-mounted robotic sensors by solving homogeneous transform equations of the form $AX = XB$. *IEEE Transactions on robotics and automation*, 5(1):16–29, 1989.

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

[46] Minas E. Spetsakis and John Aloimonos. Structure from Motion Using Line Correspondences. *The International Journal of Computer Vision*, 4:171–183, 1990.

[47] N.A. Thacker. On-line calibration of a 4-dof robot head for stereo vision. In *British Machine Vision Association meeting on active vision*, London, 1992.

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

[49] H.P. Trivedi. Semi-analytic method for estimating stereo camera geometry from matched points. *Image and Vision Computing*, 9, 1991.

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

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

[52] R.Y. Tsai and R.K. Lenz. Real time versatile robotics hamd/eye calibration using 3D machine vision. In *Proc. International Conference on Robotics and Automation*, pages 554–561, 1988.

[53] Shimon Ullman. *The Interpretation of Visual Motion*. MIT Press, 1979.

[54] T. Viéville. Auto-calibration of visual sensor parameters on a robotic head. *Image and Vision Computing*, 1993. To appear.

[55] T. Viéville and Q.T. Luong. Motion of points and lines in the uncalibrated case. Technical report, INRIA, Sept 1993.

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

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

[58] Z. Zhang. Motion and structure of four points from one motion of a stereo rig with unknown extrinsic parameters. In *Proc. International Conference on Computer Vision and Pattern Recognition*, pages 556–561, 1993.

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